

10/522037

EPO BOSTON 24 JAN 2005



Europäisches
Patentamt

European
Patent Office

Office européen
des brevets

REC'D 27 NOV 2003

WIPO PCT

Bescheinigung

Certificate

Attestation

Die angehefteten Unterlagen stimmen mit der ursprünglich eingereichten Fassung der auf dem nächsten Blatt bezeichneten europäischen Patentanmeldung überein.

The attached documents are exact copies of the European patent application described on the following page, as originally filed.

Les documents fixés à cette attestation sont conformes à la version initialement déposée de la demande de brevet européen spécifiée à la page suivante.

Patentanmeldung Nr. Patent application No. Demande de brevet n°

02291871.8

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH
RULE 17.1(a) OR (b)

Der Präsident des Europäischen Patentamts;
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

R C van Dijk

BEST AVAILABLE COPY



Anmeldung Nr.:
Application no.: 02291871.8
Demande no.:

Anmeldetag:
Date of filing: 24.07.02
Date de dépôt:

Anmelder/Applicant(s)/Demandeur(s):

Libragen
11 Boulevard Einstein
69100 Villeurbanne
FRANCE

Bezeichnung der Erfindung/Title of the invention/Titre de l'invention:
(Falls die Bezeichnung der Erfindung nicht angegeben ist, siehe Beschreibung.
If no title is shown please refer to the description.
Si aucun titre n'est indiqué se referer à la description.)

Method for the expression of unknown environmental DNA into adapted host cells

In Anspruch genommene Priorität(en) / Priority(ies) claimed /Priorité(s)
revendiquée(s)
Staat/Tag/Aktenzeichen/State>Date/File no./Pays/Date/Numéro de dépôt:

Internationale Patentklassifikation/International Patent Classification/
Classification internationale des brevets:

C12N15/00

Am Anmeldetag benannte Vertragstaaten/Contracting states designated at date of
filing/Etats contractants désignées lors du dépôt:

AT BE BG CH CY CZ DE DK EE ES FI FR GB GR IE IT LI LU MC NL PT SE SK TR

Method for the expression of unknown environmental DNA into adapted host cellsIntroduction and Background

5 The present invention relates to methods and compositions for nucleic acid production, analysis and cloning. The present invention discloses tools and methods for the production and analysis of libraries of polynucleotides, particularly metagenomic libraries, which can be used to identify novel pathways, novel enzymes and novel metabolites of interest in various areas, including pharmaceutical, cosmetic, 10 agrochemical and/or food industry.

Drug discovery process is based on two main fields, namely combinatorial chemistry and natural products. Combinatorial chemistry has shown its ability to generate huge amounts of molecules, but with limited chemical diversity. At the opposite, natural 15 products have been the most predominant source of structural and molecular diversity. However, the exploitation of this diversity is strongly hampered by their limited access, complex identification and purification processes, as well as by their production.

Microorganisms are known to synthesize a large diversity of natural compounds which 20 are already widely used in therapeutic, agriculture, food and industrial areas. However, this promising approach to the identification of new natural compounds has always been considerably limited by the principal technological bolts of isolating and *in vitro* propagating the huge diversity of bacteria. Most microorganisms living in a natural, complex environment (soil, digestive tract, sea, etc...) have not been cultivated because 25 their optimal living conditions are either unknown or difficult to reproduce. Numbers of scientific publications relate this fact and it is now assumed that less than about 1% of the total bacterial diversity (when all environments are considered together) have been isolated and cultivated (Amann *et al*, 1995).

New approaches have been developed to try to overpass the critical step of isolation, and to access directly to the huge genetic potential established by the microbial adaptation processes through their long evolution. These approaches are called "Metagenomic" 5 because they address a plurality of genomes of a whole bacterial community, without any distinction (metagenome).

Metagenomics involve direct extraction of DNAs from environmental samples and their propagation and expression into a cultivated host cell, typically a bacteria. Metagenomic 10 has been firstly developed for the identification of new bacterial phylum (Pace *et al.* 1997). This use is based on the specific cloning of genes recognized for their interest as phylogenetic markers, such as 16S rDNA genes. Further developments of Metagenomics relate to the detection and cloning of genes coding for proteins with environmental or industrial interest. These first two applications of metagenomic involve 15 a first step of selection of genes (generally using PCR) before cloning. In the case of protein production, the cloning vector used are also expression vectors, i.e., they contain regulatory sequences upstream of the cloning site causing expression of the cloned gene in a given bacterial host strain.

More recent developments of metagenomic consider the total metagenome cloned 20 without any selection and/or identification, to establish random "Metagenomic DNA libraries". This provides an access to the whole genetic potential of bacterial diversity without any "*a priori*" selection. Metagenomic DNA libraries are composed of hundreds of thousands of clones which differ from each other by the environmental DNA fragments which have been cloned. In this respect, large DNA fragments have been 25 cloned (up to 40 Kb), so as to (i) limit the number of clones which have to be analysed and (ii) to be able to recover whole biosynthetic pathways. This last point is of particular interest for bacterial metagenomic libraries since, most of the time, the biosynthetic pathways have been found to be naturally organised in a same cluster of DNA and even in the same operon in bacteria.

Except for the identification and characterisation of bacterial community at the phylogenetic or diversity levels, metagenomic libraries produced in the prior art are gene expression libraries, i.e., the environmental DNA fragments are cloned downstream of a functional promoter, to allow their expression and analysis. In this regard, WO99/45154 and WO96/34112 relate to combinatorial gene expression libraries which comprise a pool of expression constructs where each expression construct contains DNA which is operably associated with one or more regulatory regions that drive expression of genes in an appropriate organisms. Furthermore, the expression constructs used in these methods have a very limited and invariable host range. Similarly, WO 01/40497 relates to the construction and use of expression vectors which can be transferred in one chosen expression bacterial host of the *Streptomyces* genus. All these approaches are, however, very limited since they require the presence of expression signals and confer invariable or very limited host range capabilities. Furthermore, most (if not all) metagenomic DNA libraries are established in *E. coli* which is the most efficient cloning system. However, most environmental DNA are not expressed or functionally active in *E. coli*. In particular, functional analysis in *E. coli* of genes cloned from G+C rich organisms, such as Actinomyces, could be limited by the lack of adequate transcription and translation system. Also, posttranslational modification system in *E. coli* is not operative on heterologous proteins from Actinomycetes and some specific substrates for proteins activity are not present in *E. coli*.

Thus, the exploitation of the huge potential of metagenomics for the discovery of new natural compounds, pathways or genes cannot be achieved with currently existing methods. Alternative technologies and processes must be developed, to allow production of efficient libraries and functional screening in a large variety of host cell species, including *Bacillus* or *Streptomyces*, to take full account of the huge diversity of the environmental DNAs.

Summary of the Invention

The present invention discloses improved tools and methods for the production and analysis of libraries of polynucleotides, particularly metagenomic libraries, which can be
5 used to identify novel pathways, novel enzymes and novel metabolites of interest.

More particularly, the invention now proposes to keep the advantage of high efficient cloning in *E. coli* and to modify the properties of metagenomic libraries, to allow functional analysis of particular selected clones in any appropriate system, thereby
10 making possible the analysis of the huge diversity of metagenomic libraries. According to the invention, polynucleotide libraries can be produced in any convenient cloning system, such as *E. coli*, and then modified, depending on the desired selection or screening system, to adapt host range and/or properties of the library (or a portion thereof).

15

A particular object of this invention resides more specifically in a method of analysing a library of polynucleotides, said polynucleotides being contained in cloning vectors having a particular host range, the method comprising (i) selecting cloning vectors in the library which contain a polynucleotide having a particular characteristic, (ii) modifying
20 said selected cloning vectors to allow a transfer of said vectors and/or expression of the polynucleotide which they contain into a selected host cell, and (iii) analysing the polynucleotides contained in said modified vectors upon transfer of said modified vectors into said selected host cell.

25 An other object of this invention is a library of polynucleotides, wherein said library comprises a plurality of environmental DNA fragments cloned into cloning vectors, wherein said environmental DNA fragments contain a common molecular characteristic and wherein said cloning vectors are *E. coli* cloning vectors comprising a target

polynucleotide construct allowing transfer and/or expression of the environmental DNA into a selected host cell distinct from *E. coli*.

A further object of this invention is a method of producing modified libraries of 5 polynucleotides, the method comprising selecting a sub-population of clones in a first library, based on the presence or absence of a characteristic of interest, and modifying the properties of said selected clones to allow their functional analysis or expression.

The modification in the library or cloning vector is typically obtained by targeted 10 insertion of a polynucleotide construct, preferably using transposable elements, either in vitro or in vivo.

The invention is particularly suited for producing and analysing genetic diversity (metagenomic libraries), to identify and isolate new metabolites, drugs, enzymes, 15 antibiotics, etc.

Legend to the Figures

Figure 1 : Map of pP1 vector carrying transposable construct Tn<Apra> fig.1a. Excised 20 transposable construct is 1132 bps in size. It contains two mosaic ends (ME) and a gene conferring resistance to apramycin (Apra) fig.1b.

Figure 2 : pPL1 Vector (fig. 2a) that carries the conjugative transposable construct Tn<Apra-oriT> (fig. 2b). The nucleotide sequence of the transposable construct contains 25 an origin of transfer (oriT) and a gene conferring resistance to apramycin (Apra). Direction of DNA transfer at oriT is shown by an arrow.

Figure 3 : Transposable construct Tn<Apra-oriT-att-int> (fig. 3a) on vector pPAOI6 (fig. 3b). The transposable construct contains Φ C31 integrase gene and attachment DNA

sequence for site specific integration, origin of transfer and gene for selection. The orientation of genes and direction of DNA transfer are marked by arrows.

Figure 4 : This figure shows any target DNA suitable for insertion of transposable construct (fig. 4a). Insertion of conjugative and site specific integrative transposable construct Tn<Apra-oriT-int> is shown on fig. 4 b and c. Insertion of transposable construct into selective gene marker carried on original vector is shown on fig. 4b. In this event, the cloned insert is intact and can be transferred to heterologous host. Insertion of transposable construct into cloned DNA insert results in gene inactivation (fig. 4c).

Figure 5 : Complete annotated DNA sequences of clones FS3-124 (a, SEQ ID NO: 1) and FS3-135 (b, SEQ ID NO: 2).

Figure 6 : Morphological differences between *Streptomyces* transconjugant (assay). Conjugations have been performed with FS3-124 modified with transposable construct pPAOI6; (control) conjugation have been performed with pPAOI6.

Detailed Description of the Invention

20

The invention provides novel strategies, methods and products for generating and analysing combinatorial gene libraries. As indicated above, the invention discloses, particularly, methods of analysing libraries of polynucleotides, said polynucleotides being contained in cloning vectors having a particular host range, the methods comprising (i) selecting cloning vectors in the library which contain a polynucleotide having a particular characteristic, (ii) modifying said selected cloning vectors to allow a transfer of said vectors and/or expression of the polynucleotide which they contain into a selected host cell, and (iii) analysing the polynucleotides contained in said modified

vectors upon transfer of said modified vectors into said selected host cell, such as by genetic, biochemical, chemical or phenotypical approaches.

Library of polynucleotides

5

The term "library of polynucleotides" designates a complex composition comprising a plurality of polynucleotides, of various origins and structure. Typically, the library comprises a plurality of unknown polynucleotides, i.e., of polynucleotides whose sequence and/or source and/or activity is not known or characterized. In addition to such unknown (or uncharacterised) polynucleotides, the library may further include known sequences or polynucleotides. Typically, the library comprises more than 20 distinct polynucleotides, more preferably at least 50, typically at least 100, 500 or 1000. The complexity of the libraries may vary. In particular, libraries may contain more than 5000, 10 000 or 100 000 polynucleotides, of various origin, source, size, etc. Furthermore, the polynucleotides are generally cloned into cloning vectors, allowing their maintenance and propagation in suitable host cells, typically in *E. coli*. The polynucleotides in the library may be in the form of a mixture or separated from each other, in all or in part. It should be understood that some or each polynucleotide in the library may be present in various copy numbers.

20

The polynucleotides in the libraries are more preferably obtained or cloned from complex sources of nucleic acids, most preferably from environmental samples. Such libraries are also termed "metagenomic libraries" since they contain nucleic acids derived from whole genomes of complex mixtures of microorganisms.

25

The term environmental sample designates, broadly, any sample containing (a plurality of) uncharacterized (micro)organisms, particularly uncultivated (or non-cultivable) microorganisms. The sample may be obtained or derived from specific organisms, natural environments or from artificial or specifically created environments (e.g.,

industrial effluents, etc). An uncultivated (or non-cultivable) microorganism is a microorganism that has not been purposely cultured and expanded in isolated form. The sample may be obtained or derived from soil, water, mud, vegetal extract, wood, biological material, marine or estuarine sediment, industrial effluents, gas, mineral extracts, sand, natural excrements, meteorites etc. The sample may be collected from various regions or conditions, such as tropical regions, deserts, volcanic regions, forests, farms, industrial areas, household, etc.

Environmental samples usually contain various species of (uncharacterized, uncultivated) microorganisms, such as terrestrial microorganisms, marine microorganisms, salt water microorganisms, freshwater microorganisms, etc. Species of such environmental microorganisms include autotrophe or heterotrophe organisms, Eubacteria, archaebacteria, algae, protozoa, fungi, viruses, phages, parasites, etc. The microorganisms may include extremophile organisms, such as thermophiles, psychrophiles, psychrotrophes, acidophiles, halophiles, etc. More specific examples of environmental bacteria includes actinomycetes, eubacteriaes and mycobacteriaes, examples of fungi include phycomycetes, ascomycetes and basidiomycetes, etc. Other organisms include yeasts (saccharomyces, kluyveromyces, etc.) plant cells (algae, lichens, etc), corals, etc. for instance. The sample may comprise various species of such donor (uncultivated) microorganisms, as well as various amounts thereof. The environmental sample may contain, in addition, known and/or cultivable microorganisms (e.g., prokaryotic or eukaryotic), as well as nucleic acids and organic materials. The sample may also contain mammalian cells or animal cells or insect cells (arising from larvae, feces, etc.).

25

It should be understood that the present invention is not limited to any specific type of sample or environmental microorganism, but can be used to produce diversity, create nucleic acid libraries, etc., from any environmental sample comprising uncultivated microorganisms. The sample may be wet, soluble, dry, in the form of a suspension,

paste, powder, solid, etc. Preferably, the sample is dry or in solid or semi-solid state (e.g., paste, powder, mud, gel, etc.). The sample may be treated prior to nucleic acid extraction, for instance by washings, filtrating, centrifuging, diluting, drying, etc.

- 5 The term "environmental DNA" designates any DNA fragment or collection obtained from an environmental sample. Nucleic acids may be extracted / isolated from the sample according to various techniques, such as those described in WO01/81357, in WO01/40497, in Handelsman *et al.* (Chemistry & Biology 5(10), 1998, R245), Rondon *et al.* (Tibtech 17, 1999, 403 ; Applied and Environm. Microbiol. 66, 2000, 2541), Miller
10 et al (Applied and Environm. Microbiol. 65, 1999, 4715) or Frostegard *et al.* (Applied and Environm. Microbiol., 65, 1999, 5409).

In a particular embodiment of the above method, the library comprises a plurality of environmental DNA fragments. The library may also comprise other types of nucleic acids, such as environmental RNAs, for instance.
15

The polynucleotides have a size which is typically comprised between 10 and 100 kb, more preferably between 20 and 80 kb, typically between 30 and 80 kb. Although not mandatory, it is preferred that the polynucleotide fragments in the library all have similar
20 size, to produce homogenous libraries.

Cloning Vectors

As indicated, the polynucleotides are contained or cloned into cloning vectors. These
25 vectors may be of various types, including plasmids, cosmids, fosmids, episomes, artificial chromosomes, phages, viral vectors, etc. In most preferred embodiments, the cloning vectors are selected from plasmids, cosmids, phages (e.g., P1-derivatives) and BACs, even more preferably from cosmids, P1 derivatives and BACs. By using cosmids or P1 derivatives, it is possible to generate homogenous libraries, since these vectors

essentially accommodate polynucleotides having a size of approximately 40 kb and 80 Kb, respectively. Furthermore, since the invention provides that the vectors are modified after the initial cloning step, the cloning capacity of the vectors is maximized and inserts of as much as 40 and 80 kbs in length can be cloned into cosmids and P1 derivatives.

5

As indicated the cloning vector has a particular host range, i.e., the ability to replicate in a particular type of host cell. Typically, the host is a bacteria, more preferably an *E. coli* strain. Indeed, *E. coli* is so far the most convenient host cell for performing recombinant technologies. The advantage of the present invention is that the starting library can be produced in any suitable host system of choice, since the properties of the libraries will 10 be adapted later during the process.

Cloning vectors generally comprise the polynucleotide insert and genetic elements necessary and sufficient for maintenance into a competent host cell. They typically 15 contain, in addition to the polynucleotide insert, an origin of replication functional in a selected host cell as well as a marker gene for selection and screening. The cloning vector may comprise additional elements, such as promoter regions, for instance. Although cloning vectors may replicate in several different host cells, they are usually adapted to a particular host cell type and not suitable or efficient for replication or 20 maintenance in other cell types.

In a preferred embodiment, the cloning vectors of the library are *E. coli* cloning vectors, preferably cosmids , BAC or P1 derived vectors. *E. coli* cloning vectors may carry an origin of replication derived from naturally-occurring plasmids, such as ColE1, pACYC 25 and p15A, for instance. Many *E.coli* cloning vectors are commercially available and/or can be constructed using available regulatory sequences.

Screening of the cloning vectors

In step i) of the method, a first selection or screen is performed on the polynucleotide library. The screen is performed so as to identify or select clones having (or lacking) a particular, common characteristic. The selection may be carried out according to various techniques, such as by molecular screening, protein expression, functional screening, etc.

- 5 A preferred selection is performed by molecular screening. Molecular screening designates any method of identification of molecular or structural characteristics in a polynucleotide sequence. This can be made by a variety of techniques which are known per se, such as hybridisation, amplification, sequencing, etc. Preferably, molecular screening comprises the selection of clones in a library which contain, in their sequence,
- 10 a particular sequence or region or motif, said sequence or region or motif being characteristic of a particular type of activity or gene (enzyme, biosynthetic pathways, etc.).

In a first variant, the selection is made by contacting the cloning vectors in the library with a particular nucleic acid probe (or set of probes) containing a sequence which is characteristic of a selected activity or function (a consensus sequence, a particular motif, etc.). The cloning vectors in the library which hybridise to the probe (or set of probes) are then selected.

- 20 In a second variant, the selection is made by contacting the cloning vectors in the library with a particular pair of nucleic acid primers specific for a sequence which is characteristic of a selected activity or function (a consensus sequence, a particular motif, etc.), and a PCR amplification reaction is performed. The cloning vectors in the library which lead to a positive amplification product are then selected.

25

In this regard, the present application provides new primers designed in conserved motives of the β -keto acyl synthase gene, which are particularly useful for screening polynucleotides containing putative PKS genes or domains. These primers have the following degenerated sequence :

Sense primer : 5'- GGSCCSKCSSTSDCSRTSGAYACSGC -3' (SEQ ID NO: 3)

Antisense primer : 5'- GCBSSRYYTCDATGGRTCSCC -3' (SEQ ID NO: 4)

wherein :

R is A or G

5 S is G or C

Y is C or T

K is G or T

D is A or G or T, and

B is C or G or T.

10

A particular object of this invention is a polynucleotide primer having one of the above sequences, typically a mixture of different polynucleotide primers having a sequence corresponding to one of the above degenerated sequences. A particular object of this invention also resides in a pair of primers each having one of the above sequences.

15

Once particular clones have been selected, the analysis of their polynucleotides needs to be confirmed and/or validated, and/or their polynucleotides can be used to produce novel compounds or metabolites.

20

The invention now enables such further analysis and uses, by allowing a modification of the cloning vectors that is specific and adaptable by the skilled person, depending on the activity which is sought. In particular, it is possible to confer properties such as specific expression or a novel, specific host range to the selected cloning vectors, to assess their activity, as disclosed below.

25

Modification of the cloning vectors

After high efficiency cloning using most convenient cloning vectors such as BACs or cosmids propagated into *E. coli*, and after the identification, selection and/or

characterisation of cloned DNA fragments, the invention now allows to modify specifically the cloning vectors to transfer, maintain, express and/or over-express the selected polynucleotides into any selected host expression system, which is suitable to assess the selected activity or property. Such selected hosts may be native or 5 heterologous host cells, and include, but are not limited to, for example *Streptomyces*, *Nocardia*, *Bacillus*, fungi, yeasts, etc.

The selected cloning vectors of the library may be modified according to various techniques. The modification is typically a genetic modification, comprising the 10 introduction of particular genetic sequences into the structure of the cloning vector, in addition to or in replacement of sequences contained in said vector. It is highly preferred to use specific or targeted (or oriented) techniques to improve the efficacy of the method. By "specific" is meant that the modification occurs at a pre-determined location in the cloning vector, through site-specific mechanisms. By "targeted" is meant that the 15 modification occurs in a controlled way, so as not to alter the polynucleotide insert contained in the vector in a non-desirable way.

In a preferred embodiment, the selected vectors are modified by insertion, into the vector, of a target polynucleotide construct which contains genetic elements conferring 20 the selected property(ies) to the cloning vector.

The target polynucleotide construct typically comprises the genetic elements necessary to transfer, propagate, maintain , express or overexpress the cloned polynucleotide into a chosen (bacterial) host expression system. Said genetic elements may include particular 25 origin(s) of replication, particular origin(s) of transfer, particular integrase(s), transcriptional promoter(s) or silencer(s), either alone or in combination(s).

In a first, preferred variant, the target polynucleotide construct comprises a genetic element allowing transfer of the vector into a selected host cell.

- Natural DNA transfer mechanisms between donor and recipient strains is known under the term conjugation or conjugative transfer. Conjugative transfer can occur between different strains of the same species as well as between strains of different species. Many naturally occurring plasmids carry so called *tra* genes, which are involved in and mediate conjugative transfer. The DNA transfer starts at specific DNA structures, known as an origin of transfer or "ori T". The presence of such an oriT in a vector allows said vector to be transferred into a desired host cell.
- 10 In a particular, preferred embodiment, the target polynucleotide construct comprises an origin of transfer functional in the selected host cell, more particularly in a *Streptomyces* or *Bacillus* host cell.

15 The structure of various oriT has been reported in the art (Guiney *et al.* 1983; Zechner *et al.* 2000). In a specific embodiment, the origin of transfer is selected (or derived) from RP4, pTiC58, F, RSF1010, ColE1 and R6K(α).

A specific example of an oriT which can be used in the present invention derives from plasmid RP4 and has or comprises all or a functional part of the following sequence
20 (SEQ ID NO: 5) :

```

gatctGTGATGTACTTCACCAGCTCCCGAAGTCGCTCTTGATTGGAGCGCATGGG
GACGTGCTTGGCAATCACGCCACCCCCCGGCCGTTAGCGGCTAAAAAAAGTCATG
GCTCTGCCCTCGGGCGGACCACGCCATCATGACCTTGCCAAGCTCGTCCTGCTTCT
25 CTTCGATCTCGCCAGCAGGGCGAGGATCGTGGCATACCGAACCGCGCCGTGCGC
GGGTCGTCGGTGAGCCAGAGTTTCAGCAGGCCCGCCAGGCGGCCAGGTGCGCCATT
GATGCGGGCCAGCTCGCGACGTGCTCATAGTCCACGACGCCGTGATTGTAGCC
CTGGCCGACGGCCAGCAGGTAGGCCACAGGCTCATGCCGGCCGCCGCGCCTTTC
CTCAATCGCTTCTCGTCTGGAAAGGCAGTACACCTGATAGGTGGGCTGCCCTT
30 CCTGGTTGGCTTGGTTCATCAGCCATCCGCTGCCCTCATCTGTTACGCCGGCGTA
GCCGGCCAGGCTCGCAGAGCAGGATTCCCGTGAGCACCGCCAGGTGCGAATAAGG
GACAGTGAAGAAGGAACACCCGCTCGCGGGTGGGCCTACTTCACCTATCCTGCCCG
GCTGACGCCGTTGGATACACCAAGGAAAGTCTACACGAACCCCTTGGCAAAATCCT

```

GTATATCGTGCAGAAAAGGATGGATATAACGAAAAAATCGCTATAATGACCCGAA
GCAGGGTTATGCAGCGGAAAAGATCCGTCGGATCT

The term "functional part" designates any fragment or variants of the above sequence
5 which retain the capacity to cause conjugative transfer. Such fragments typically comprise at least 80%, preferably at least 85% or 90% of the above sequence. Variants may include one or several mutations, substitutions, deletions or additions of one or several bases.

10 In an other particular variant, the target polynucleotide construct comprises a genetic element allowing integration of the vector into the genome of the selected host cell.

A donor DNA is permanently or stably maintained and expressed in a selected recipient cell if it is integrated into the recipient cell's genome or if it contains elements that allow
15 autonomous replication in said cell.

Integration is a preferred way of ensuring stable expression. Integration can be obtained by physical recombination. Recombination can be homologous, e.g., between two homologous DNA sequences, or illegitimate, where recombination occurs between two
20 non-homologous DNAs. As a particular example, integration of donor DNA into the chromosome of the recipient can be mediated by host recombination repair system or by site-specific recombination. Another well-studied process that can transfer and integrate genes is transduction by bacterial viruses, such as λ and ϕ C31. In a phage-infected bacterial cell, fragments of the host DNA are occasionally packaged into phage particles
25 and can then be transferred to a recipient cell. Integration into the recipient cell's genome is caused by an integrase.

In a specific embodiment, the target polynucleotide construct comprises a nucleic acid encoding an integrase functional in the selected host cell. More preferably, the integrase
30 is selected from λ and ϕ C31 integrases. In a specific embodiment, the polynucleotide

construct comprises a nucleic acid encoding an integrase having or comprising all or a functional part of the following sequence of the φC31 integrase (SEQ ID NO: 6) :

```

5   AGATCTCCGTACTGACGGACACACCGAAGCCCCGGCAACCTCAGCGGATGC
    CCCGGGGCTTCACGTTTCCCAGGTCAAAGCGGTTTCGGGAGTAGTGCCCCAACT
    GGGGTAACCTTGAGTTCTCTCAGTTGGGGCGTAGGGTCGCCACATGACACAAG
    GGGTTGTGACCGGGGTGGACACGTACGCCGGTGCTTACGACCGTCAGTCGCGAG
    CGCGAGAACATTGAGCGCAGCAAGCCCAGCGACACAGCGTAGCGCCAACGAAGACA
    AGGCGGCCGACCTTCAGCGCGAAGTCGAGCGCAGCGGGGCCGTTCAAGGTTCGTC
    10  GGGCATTTCAGCGAACGCCCCGGCACGTGGCTTGGGACGGCGAGCGCCCGGA
    GTTCGAACGCATCCTGAACGAATGCCGCCGGCGCTCAACATGATCATTGTCTA
    TGACGTGTGGCGCTTCGCGCCTGAAGGTCAATGGACGCGATTCCGATTGTCTCGGA
    ATTGCTGCCCTGGCGTGACGATTGTTCCACTCAGGAAGGCGTCTCCGGCAGGG
    AACGTCATGGACCTGATTCACCGATTATGCGGCTCGACCGCTCGCACAAAGAAC
    15  TTCGCTGAAGTCGGCGAACGATTCTGACACGAAGAACCTTCAGCGCGAATTGGGCG
    GGTACGTGGCGGGAACGGCGCCTACGGCTTCAGGCTTGGAGACGAAGGAG
    ATCACGCGAACGGCGAACGGTCAATGTCGTCAACACAAGCTTGCCTCGACCGACC
    ACTCCCCCTACCGGACCCCTCGAGTTGAGCCCAGCTAACCGGTGGTGGTGGCGT
    GAGATCAAGACGCACAAACACCTCCCTCAAGCCGGCAGTCAGCCGCCATTCA
    20  CCCGGGCAGCATCACGGGGTTGTAAGCGCATGGACGCTGACGCCGTGCCGACCC
    GGGCGAGACGATTGGGAAGAACGACCGCTCAAGGCCCTGGACCCGGCAACCGTT
    ATGCGAATCCTCGGGACCCCGTATTGCGGGCTTCGCGCTGAGGTGATCTACAAG
    AAGAAGCCGGACGGCACGCCGACCGAACGAAAGATTGAGGGTACCGCATTCAAGCGCGA
    CCCGATCACGCTCCGGCCGGTCAGCTTGGATTGCGGACCGATCATCGAGCCCGCTGA
    25  GTGGTATGAGCTTCAGGCGTGGITGGACGGCAGGGGGCGCGCAAGGGGTTTCCC
    GGGGGCAAGCCATTCTGCGGCCATTGACAAGCTGTACTGCGAGTGTGGCGCCGTC
    ATGACTTCGAAGCGGGGGAAAGAACGATCAAGGACTCTTACCGCTGCCGTGCGCCG
    GAAGGTGGTCGACCCGTCCGCACCTGGGCAGCACGAAGGCACGTGCAACGTCAAG
    TGGCGGCACTCGACAAGTTGTTGGGAACGCATCTCAACAAAGATCAGGCACGCC
    30  GAAGGCACGAAGAGACGTTGGCGCTTGTGGGAAGCCGCCGACGCTTGGCAA
    GCTCACTGAGGCCTGAGAACGAGAGCGCGAACGGCGAACCTTGTGGAGCGCG
    CCGACGCCCTGAACGCCCTGAAGAGCTGTACGAAGACCGCGCGCAGGCGCGTAC
    GACGGACCCGTTGGCAGGAAGCACTCCGAAGCAACAGGCAGCGCTGACGCTCCG
    GCAGCAAGGGGGCGGAAGAGCGGCTGCCAACCTGAAGCCCGAACGCCCGAAG
    35  CTTCCCTTGACCAATGGTTCCCCGAAGACGCCGACGCTGACCCGACCGCCCTAAG
    TCGTGGTGGGGCGCGCTCAGTAGACGACAAGCGCGTGTGCTCGGGCTCTCGTA
    GACAAGATCGTTGTCACGAAGTCGACTACGGCAGGGGGCAGGGAACGCCCATCGA
    GAAGCGCGCTTCGATCACGTGGCGAACGCCCGACCGACGACGACAAGACGACG
    CCCAGGACGGCACGGAAGACGTAGCGCGTAGCGAGACACCCG

```

40

In a more preferred variant, the target polynucleotide construct comprises genetic elements allowing transfer of the cloning vector into the selected host cell and integration of the cloning vector or a portion thereof into the genome of the selected host

cell. Most preferred polynucleotide constructs comprises an oriT and a nucleic acid encoding an integrase.

In an other variant, the target polynucleotide construct comprises an origin of replication 5 specific for or functional in the selected host cell. The origin of replication may be selected (or derived), for instance, from pAMβ1, pSa, 2μm circle, pSam2, pSG1, pIJ101, SCP2, pA387 and artificial chromosomes.

In an other variant, the target polynucleotide construct comprises a transcriptional promoter functional in the selected host cell. As indicated above, in a particular variant, 10 the invention allows to modify the cloning vector to enable expression or over-expression of the cloned polynucleotides in the selected host. The expression of genes is driven mainly by transcriptional promoters, which initiate gene transcription. The type of promoter to be used in the present invention can be selected by the skilled person, 15 depending on the selected host cell and type of expression needed. Promoters may be ubiquitous or cell-specific, regulated or constitutive, weak or strong. They may be of various origins, including promoters isolated from viruses, phages, plant cells, bacterial genes, mammalian genes, etc., or they may be artificial or chimeric. Typical examples of promoters include T7, T4, LacZ, trp, ara, SV40, tac, λPL, GAL, AOX, hsp-70, etc.

20 The target polynucleotide construct is typically a DNA molecule, although RNAs may also be used as starting material. It is typically a double-stranded DNA. The target polynucleotide construct may be produced by conventional recombinant DNA techniques, including DNA synthesis, cloning, ligation, restriction digestion, etc. and a 25 combination thereof.

The target polynucleotide construct is preferably engineered so as to be inserted in a region of the vector distinct from the polynucleotide. Indeed, it is important that the integrity of the polynucleotides is preserved. Directed insertion may be accomplished in

a variety of ways, including site-specific insertion using particular enzymatic systems (Cre/Lox, FLP, etc.), homologous recombination with particular target sequences present in the vector, or by the use of transposons or transposable elements and appropriate selection means.

5

In a particular, preferred embodiment, the target polynucleotide construct is contained in a transposable nucleic acid construct. Indeed, in a preferred variant, the methods of the present invention use transposable elements to alter the properties of the cloning vectors, and allow their transfer, maintenance, expression or over-expression in a selected host cell.

10

Transposable nucleic acid constructs are derived from transposons, which are genetic elements capable of moving from one genetic loci to another. Two main classes of transposons have been identified in bacteria. The most simple transposons comprise an 15 insertion sequence that carries only elements of transposition. These elements are two inverted DNA repeats and a gene that codes for a protein called transposase. The transposase catalyses the excision and integration of the transposon. It has been shown that the excision and integration reaction can be catalysed in trans by a transposase, which can be provided in vivo or in vitro in purified form or expressed from a different 20 construct. More complex transposons carry more insertion sequences and additional genes that are not involved in transposition.

20

Transposable nucleic acid constructs of this invention thus typically comprise, flanked by two inverted repeats, the target polynucleotide construct and, more preferably, a 25 marker gene. In the presence of a transposase, these transposable nucleic acid constructs can integrate into a cloning vector in vivo or in vitro, thereby providing for targeted polynucleotide insertion. Alternatively, such nucleic acid constructs can be used for targeted integration, in the absence of a transposase, in particular strains such as hypermutator strains. Such transposable nucleic acid constructs also represent a

particular object of the present application. In this regard, in a more preferred embodiment, the invention also relates to a transposable nucleic acid construct, wherein said construct comprises an origin of transfer flanked by two inverted repeats. Specific examples of such construct are transposons pPL1 and pPAOI6, as disclosed in the 5 experimental section. The transposable nucleic acid construct may further comprise an integrase gene and/or a marker gene.

The inverted repeat nucleic acid sequences may be derived from the sequence of various transposons, or artificially created. In particular, transposable elements can be generated 10 using inverted repeats obtained from transposons or transposable elements such as Tn5, Tn21, miniTn5, T7, T10, Tn917, miniTn400, etc. Preferably, the sequences derive from transposon Tn5. In a specific embodiment, they comprise all or a functional part of the following sequences :

- 15 - left arm of pPAOI6 transposon (SEQ ID NO: 7)
CTGTCTCTTATACACATCTCAACCATCATCGATGAATTTCAGGGTGTCTCGCATA
TTGGCTCGAATTGAGCTCGGTACCC
- right arm of transposon pPAOI6 (SEQ ID NO: 8)
20 GATCCTCTAGAGTCGACCTGCAGGCATGCAAGCTGCCAACGACTACGCACTAGCCA
ACAAGAGCTTCAGGGTTGAGATGTGTATAAGAGACAG

The marker gene may be any nucleic acid encoding a molecule whose presence in a cell can be detected or visualized. Typical marker genes encode proteins conferring 25 resistance to antibiotics, such as apramycin, chloramphenicol, ampicilin, kanamycin, spectinamycin, thiostrepton, etc. Other types of markers confer auxotrophy or produce a label (e.g., galactosidase, GFP, luciferase, etc).

In a specific embodiment, the cloning vector in the library comprises a first marker gene 30 and the modification step ii) comprises:

- . contacting in vitro, in the presence of a transposase, the selected cloning vectors with a transposon comprising, flanked by two inverted repeats, the target polynucleotide construct and a second marker gene distinct from the first marker gene, and
- . selecting the cloning vectors which have acquired the second marker gene and which
5 have lost the first marker gene.

The double selection ensures that the target polynucleotide construct has been inserted at a site within the marker gene present in the cloning vector, i.e., outside of the polynucleotide insert.

- 10 It should be understood that the modification may be accomplished in various other ways, particularly by incorporating a sequence coding for the transposase directly into the transposable element or into another expression unit. The presence and expression of the transposase can be regulated by inductive promoter or termosensitive replicative units. Also, transposition can be carried out by *in vitro* process.
15

Analysis of the polynucleotides

- In step (iii) of the process, the polynucleotides may be analysed by various methods,
20 including by genetic, biochemical, chemical or phenotypical approaches, which are well known per se in the art. Analysis occurs upon transfer and, optionally, expression of the polynucleotides into the selected host cell.

- In this regard, the modified cloning vectors can be transferred into the selected host cell
25 by a variety of techniques known in the art, including by transformation, electroporation, transfection, protoplast fusion, conjugative transfer, etc. In a preferred embodiment, the target polynucleotide construct comprises an oriT and the modified vectors are transferred into the selected host cells by conjugative transfer. In this embodiment, the

cloning vector and the selected host cells are co-cultivated and the recombinant host cells are selected and isolated.

5 The selected host cell may be any type of cell or microorganism, including, without limitation, *Streptomyces*, *E. coli*, *Salmonella*, *Bacillus*, *Yeast*, *fungi*, etc.

One of the objectives of the invention is to be able to analyse environmental DNAs of unknown cellular origin into different host expression systems. In order to analyse the potentiality of the DNAs at the transcription and/or translation levels and to have much 10 more probabilities to have a DNA expression, it is important to have the possibility to test different host expression systems.

Insertion of foreign DNA into host expression systems like *Streptomyces* can produce, for instance, an increase in doubling time, morphological modifications, pigments 15 production, etc. which can be related either directly to the expression of foreign DNA or by combinatorial biology of the foreign DNA and the biology of the expression host systems. The new characters can be analysed by all techniques known in the art such genetic, biochemical, chemical, phenotypic approaches etc...

20 In a preferred, specific embodiment, the invention relates to a method for the identification or cloning of polynucleotides encoding a selected phenotype, the method comprising (i) cloning environmental DNA fragments into *E. coli* cloning vectors to produce a metagenomic library, (ii) identifying or selecting cloning vectors in said library which contain DNA fragments having a particular characteristic of interest, (iii) 25 modifying the identified or selected cloning vectors into shuttle or expression vectors for transfer or expression in a selected host cell, (iv) transferring or expressing the modified cloning vectors into said selected host cell and (v) identifying or cloning the DNA fragments contained in said modified cloning vectors which encode said selected phenotype in said selected host cell.

By applying the above method, new polynucleotide sequences have been identified, cloned and characterized, which produce new phenotypes in bacteria. These polynucleotides contain the sequence of PKS genes and encode polypeptides involved in biosynthetic pathways. The sequence of these polynucleotides is provided in Figure 5. The invention also relates to any polynucleotide sequence comprising all or part of these sequences (i.e., SEQ ID NOs: 1 or 2), their complementary strand, or a functional variant thereof. A part of the above sequences includes, preferably, at least 20 consecutive bases, more preferably at least 50 consecutive bases thereof, even more preferably a coding sequence (e.g., an ORF). In this respect, SEQ ID NOs: 1 and 2 comprise several novel open reading frames encoding novel polypeptides involved in biosynthetic pathways. These coding sequences are identified in Figures 5a and 5b.

In a specific embodiment, the invention relates to a polynucleotide sequence comprising a sequence selected from nucleotides 76 - 1134 ; 1096 - 2430 ; 1178 - 1624 ; 2506 - 3567 ; 2906 - 4222 ; 4092 - 5321 ; 6337 - 8502 ; 8181 - 9530 ; 9531 - 10721 ; 10504 - 11274 ; 12874 - 13689 ; 14195 - 15976 ; 15427 - 16512 ; 15579 - 16253 ; 16505 - 17656 ; 17657 - 18697 ; 18615 - 19304 ; 19301 - 20596 ; 20535 - 21476 ; 22025 - 22951 ; 23155 - 26523 ; 26409 - 34433 ; 34418 - 37500 and 35359 - 37500 of SEQ ID NO: 1 (figure 5a) or a complementary strand thereof.

In an other specific embodiment, the invention relates to a polynucleotide sequence comprising a sequence selected from nucleotides 3 - 914, 924 - 2168 ; 2207 - 3190 ; 3373 - 4455 ; 4546 - 4959 ; 5176 - 6192 ; 6331 - 14043 ; 14275 - 15408 ; 15436 - 16245 ; 16287 - 17384 ; 17427 - 18158 ; 18248 - 18847 ; 18952 - 20346 ; 20442 - 21167 ; 21164 - 24301 ; 24351 - 27023 ; 27806 - 29686 ; 29535 - 30872 ; 30848 - 32647 ; 32574 - 35555 ; 35533 - 36598 and 36516 - 37400 of SEQ ID NO: 2 (figure 5b) or a complementary strand thereof.

The invention also relates to any vector comprising these polynucleotide sequences. These sequences may be DNA or RNA, preferably DNA, even more preferably double-stranded DNA. The invention also relates to a polypeptide encoded by a polynucleotide sequence as defined above.

5

An other object of this invention resides in a library of polynucleotides, wherein said library comprises a plurality of environmental DNA fragments cloned into cloning vectors, wherein said environmental DNA fragments contain a common molecular characteristic and wherein said cloning vectors are *E. coli* cloning vectors comprising a target polynucleotide construct allowing transfer or expression of the environmental DNA into a selected host cell distinct from *E. coli*.

10 The sub-DNA-libraries should have either desired genetic characteristics based on high or low GC content, DNA encoded for a desired enzymatic activity, part or full biosynthetic pathways for metabolites etc., or specific origin such as soil fractions, animal organs, sub fraction of a microorganism community etc. The invention also allows to produce conjugative vector with desired characteristics in accordance with the characteristics of the pre-identified sub-DNA-libraries and functional analysis of mutants in heterologous hosts. It can also be used, without limitation, for the production of mutants by random or directed mutagenesis, for DNA sequencing, genes or biosynthetic pathways knock-out by deletion or insertion or to confer transfer capabilities for expression, co-expression, over-expression or modification of biosynthetic pathways.

15 Further aspects and advantages of the invention will be disclosed in the following examples, which should be regarded as illustrative and not limiting the scope of this application.

Experimental Section

In this work, we constructed a fosmid library in *E. coli* from DNA prepared directly from soil. We screened the library for presence of biosynthetic pathways. We developed genetic tools for functional genomics that allow gene inactivation and horizontal gene transfer from *E. coli* to *Streptomyces*.

5

The cloning vectors in the library contain ColE1 replicon for propagation in *E. coli*. Transposable elements based on Tn 5 transposon were produced and used for *in vitro* (or *in vivo*) modification of selected cloning vectors. Integrated transposable elements contain gene for resistance to apramicyne. Conjugative derivatives were constructed by 10 incorporating origin of transfer from plasmid RP4. A conjugative and site specific integrative transposon was also constructed comprising the integrase gene from ϕ C31 phage, including attP attachment site. Conjugal transfer was demonstrated from an appropriate *E. coli* donor cell to another *E. coli* or *Streptomyces lividans* recipient cell. Constructed transposon was tested for inactivation of the genes cloned into fosmids. 15 Obtained mutants can be used for direct sequencing by adequate primers and transferred by conjugation into *Streptomyces lividans*. Transposable elements thus represent very useful tools for functional analysis of a large DNA libraries cloned into BAC, PAC fosmids or other cloning vectors in which cloned inserts must be transferred into heterologous host.

20 **Materials and methods**

Bacterial strains, plasmids and growth conditions.

E. coli DH10B (F- mcrA delta(mrr-hsdRMS-mcrBC) phi80dlacZ deltaM15 delta lacX74 deoR recA1 endA1 araD139 delta (ara, leu)7697 galU galK lambda- rpsL nupG), strain (Epicentre) was used for fosmid and plasmid transformation and DNA amplification.

25 Unless specifically described, all DNA manipulations were performed according to Sambrook, J., et al, (1989).

Soil DNA extraction and DNA libraries construction.

Total bacterial community DNA was extracted and large DNA libraries have been constructed into fosmids according to the method described in WO 01/81357.

Fosmid DNA extraction and Purification

5

Fosmids DNA containing soil library were extracted from pools of 96 clones. Culture of recombinant clones were performed in Deep-Well 96 and 48, respectively in 1 ml and 2 ml of LB media containing $12,5 \mu\text{g ml}^{-1}$ of chloramphenicol. Cultures were grown at 37°C with shaking at 250 RPM during 22 hours. DNA extraction was done by using the

10 Nucleobond PC100 extraction kit (Macherey Nagel).

PCR screening for the detection of PKS genes

Primers design

Degenerate PCR primers sets were designed to specifically amplify PKS nucleic sequences. Multiple sequence alignment of PKS domains revealed highly conserved motives, in particular in the β -keto acyl synthase domain. Primers Lib1F and Lib2R were designed in conserved motives of the β -keto acyl synthase gene. Lib1F (sense primer, 5'- GGSCCSKCSSTSDCSRTSGAYACSGC -3') and Lib2R (antisense primer, 5'- GCBBSRYYTCDATSGGRTCSCC -3') were deduced from β -keto acyl synthase peptide sequences GP(AS)(LV)(AST)(IV)DTAC and GDPIE(TVA)(RAQ)A, respectively. The specific fragment amplified with Lib1F / Lib2R was approximately about 465 bp (corresponding to 155 amino acids). Specificity and efficiency of the PCR systems were validated by testing on positive DNA controls (i.e. genomic DNA from type I PKS producing strain such as *Bacillus subtilis*, *Streptomyces lividans*, *Streptomyces ambofaciens* and *Ralstonia solanacearum*) and negative DNA controls (genomic DNA from strains which are known to do not contain PKS genes). Furthermore, DNA extracted from soil samples were tested to calibrate PCR techniques.

PCR conditions

PCR conditions were optimised, in particular for concentrations of DMSO, MgCl₂, Primers and DNA quantities. For PCR using microorganism genomic DNA and soil DNA as template (50 to 200 ng), the PCR mix (50µl) contained 250µM of dNTP, 5 mM MgCl₂ final, 2,5% DMSO, 1X PCR buffer, 0,75 µM of each primer and 2,5 U of Taq DNA polymerase (Sigma) and sterile distilled water. For PCR using fosmid pooled DNA as template (100 to 500 ng), the PCR mix (50µl) contained 250µM of dNTP, 5 mM MgCl₂ final, 5% DMSO, 1X PCR buffer, 0,75 µM of each primer and 2,5 U of Taq DNA polymerase (Sigma) and sterile distilled water. For identification of positives clones in 96 microtiter plates, 25 µl of each bacteria culture were used as template and

5 PCR conditions were the same as above. Thermocycling program was : a denaturation step at 96°C for 5 minutes; then 1 minute at 96°C, 65°C for 1 minute, 72°C for 1 minute. The first 7 cycles, the annealing temperature was lowered 1°C per cycle until 58°C was reached. A subsequent 40 cycles were carried out with the annealing temperature at 10 58°C. A final extension step was at 72°C for 7 minutes. For identification of positives clones in 96 microtiter plates, the first denaturation step of 96°C was during 8 minutes.

15 The other steps were the same as described above. PCR reactions were performed with a PTC 200 thermocycler (MJ Research).

PCR products analysis

20 PCR products of about 465 bp were purified on agarose gel with gel extraction Kit (Qiagen) according to the manufacturer recommendations. First approach consisted in subcloning PCR products using the Topo PCR II kit (Invitrogen). Recombinant Plasmids were extracted using QIAprep plasmid extraction Kit (Qiagen) and sequenced with Forward and Reverse M13 primers with CEQ 2000 automated sequencer (Beckman Coulter). Second approach consisted in direct sequencing of PCR products. Sequencing data were compared with nucleic and proteic genbank database using BLAST program.

25

Sequencing of the identified fosmid insert DNA and sequence analysis

Fosmids inserts were sequenced using either a transposon-mediated and by shotgun subcloning approach. Transposition was realized by using (Transposition Kit) commercialized by Epicentre according to the manufacturer. For shotgun subcloning, transformants were grown for 16 hours at 37°C. Fosmid extraction was done by using 5 the Nucleobond PC100 extraction kit (Macherey Nagel). DNA was partially restricted with *Sau3A* and sized on standard gel electrophoresis for fragments ranging from 1 to 3 Kbs and cloned into Bluescript vector according to Sambrook *et al.* (1989). Sequence analysis was performed with the identification of ORFs by using Frameplot of the GC3. Each identified ORF was compared to gene databases by using BLAST 10 program. PKS domains were determined by aligning obtained sequence versus already described PKS domains from domain databases.

Sequencing

Sequencing reactions were performed with 1 µg of DNA and 3.6 pmol of primer, using CEQ 2000 Dye Terminator Cycle Sequencing kit (Beckman Coulter) under conditions 15 proposed by supplier. Ten µL of reaction products were precipitated using 4 µL of solution containing 1.5 M NaOAc, 50 mM EDTA and 60 µL cold 95% ethanol/dH₂O from -20 °C. The pellet was washed 2 times with 200 µL 70% ethanol/dH₂O, vacuum dried and dissolved in 40 µL sample loading solution (supplied in kit). Sequencing reactions were run on an CEQ 2000 sequencer (Beckman Coulter).

20

Plasmids construction and validation

Plasmid pP1 was constructed as follows. A 941 bp DNA fragment containing native promoter region and AA(3)IV gene was amplified by polymerase chain reaction (PCR) using primers AmF (d-CCCTAACGATCTGGTTCATGTGCAGCTCCATC, SEQ ID 25 NO: 9) and AmR (d-TAGTACCCGGGGATCCAACGTCATCTCGTTCTCC, SEQ ID NO: 10). One hundred microlitre reaction were performed containing 0.1µM each of primers, 1 X Vent DNA polymerase buffer (NEB), 0.2 µM of each deoxyribonucleoside triphosphate (dNTP), 50 ng of the DNA template and 2U of Vent DNA polymerase

(NEB). PCR mixture was heated for 4 min at 94 °C in a PTC-200 thermocycler (Peltier) and cycled 25 x at 94 °C for 60 sec, 59 °C for 30 sec and at 72 °C for 70 sec. The final extension was performed at 72 °C for 7 min.

PCR product was purified using GFX DNA purification kit (Amersham), then digested 5 by *Bgl* II and *Sma* I restriction enzymes. A 941 bp *Bgl*II/*Sma*I fragment was inserted into the *Bam* HI, *Sma* I sites of pMOD plasmid (EPICENTRE). DH10B *E. coli* was transformed with pP1 and subjected to apramicyne selection on LB agar plates. Six colonies surviving on apramicyne selection were grown in liquid LB media and final pPL1 candidates were thoroughly checked via PCR and restriction mapping.

10

To construct conjugative plasmid pPL1, 750 bp oriT DNA region from plasmid RP4 were amplified via PCR using primers oriTF (d-GCGGTAGATCTGTGATGTACTTCACCAAGCTCC, SEQ ID NO: 11) and oriTR (TAGTACCCGGGGATCCGACGGATCTTTCCGCTGCAT, SEQ ID NO: 12). PCR conditions were as above. Amplified DNA was digested using *Bgl* II and *Sma* I restriction enzymes. *Bgl*II/*Sma*I DNA fragment was subjected to purification after gel electrophoresis on 0.7% agarose. Purified fragment was ligated into pP1 plasmid digested 15 by *Bam* HI and *Sma* I restriction enzymes.

20

φ C31 integrase gene and attachment site (attP) was amplified via PCR using primers Fint (d-AACAAAGATCTCCGTACTGACGGACACACCG, SEQ ID NO: 13) and PJ (d-CGGGTGTCTCGCATGCCGCT, SEQ ID NO: 14). Amplified DNA fragment was purified by GFX kit (Amersham) and phosphorilated using T4 polynucleotide kinase (NEB) under conditions recommended by the enzyme manufacturer. Phosphorilated 25 DNA fragment was cloned in to pPL1 vector opened with *Sma*I restriction enzyme (NEB) and dephosphorilated by calf alkaline phosphatase (NEB). DH10B *E. coli* was transformed with ligation mixture using Bio-Rad Pulsing apparatus and protocols provided by Bio-Rad. Twelve transformants were analyzed by PCR for the presence of

integrase gene. Orientation of integrase gene was verified by restriction analysis using *Bgl* II and *Eco*RI restriction enzymes. Resulting plasmid was named pPAOI6.

To construct pPAOI6-A plasmid, pPAOI6 plasmid was digested with *Eco*RI and *Bgl*II restriction enzymes followed by digestion by Bean mung nuclease (NEB). Linearised

5 plasmid was self ligated and transformed in to DH10B cells.

Plasmid preparation of fosmid DNA

Fosmid and BAC DNA for sequencing was prepared by using the Nucleobond AX kit (Macherey-Nagel), following protocol for BACs, Cosmid as specified by manufacturer.

Mutagenesis

10 Transposon Tn-pPAOI6 was prepared by digestion of pPAOI6 plasmid using *Pvu*II restriction enzymes, followed separation on agarose gel and purification of fragment containing transposon from gel using Qiagen kit. The same molar ratio of transposon and corresponding fosmid was used for mutagenesis in vitro using Tn5 transposase (Epicentre) and conditions specified by manufacturer. We transformed aliquot of the

15 transposed mixture by electroporation into competent DH10B *E. coli* strain.

Conjugation *E. coli*- *Streptomyces lividans* TK24

Conjugation experiments were done using 6×10^6 *E. coli* S17.1 cells containing conjugative plasmids or fosmids. The *E. coli* cells were grown in LB media with adequate antibiotic. The cells were collected by centrifugation, washed two times using same volume of LB media and concentrated to 10^8 cells/ml and overlaid on LB plates containing 2×10^6 pregerminated *Streptomyces lividans* TK24 spores. The cells mixture were grown over night at 30°C and *E. coli* cells were washed three times using 2 ml of LB media. The plates were overlayed using top agar containing NAL (nalidixic acid) and the appropriate antibiotic. Plates were incubated for 4 days at 30°C and transformant streptomyces colonies were isolated on HT medium (Pridham *et al.* 1957) containing NAL and the same appropriate antibiotic.

Results

Construction of a Transposon Tn <Apra>.

E. coli aminoglycoside -(3)- acetyl transferase IV gene (aa(3)IV) was amplified by PCR and cloned in pMOD vector (Epicentre). Advantage of this selective marker allow positive selection in *E. coli* and in *Streptomyces lividans*. Transposon can be used for insertional inactivation in vitro using purified transposase Tn5 (Epicentre). The structure of the pP1 constructed vector and transposon was shown on Figure 1a,1b.

Construction of a conjugative plasmid-transposon.

Conjugative vector, transposon was constructed by cloning origin of transfer from plasmid RP4 into pP1 vector producing pPL1 vector (Figure 2a, 2b). The origin of transfer was cloned in such orientation that the selective aa(3)IV gene is the last transferred during conjugation. PPL1 vector was introduced in to specific *E. coli* S17.1 strain that carry RP4 plasmid integrated in to chromosome. In conjugation experiment between donor strain S17.1 carrying pPL1 plasmid and DH10 *E. coli* receptor we obtained DH10B strain carrying pPL1 plasmid. This data shows that cloned oriT fragment is functional in the pPL1 plasmid. Plasmid pPL1 can be used for DNA cloning, gene inactivation by homologous recombination. Cloned genes or part of the gene cloned could be then transferred by conjugation in to another host. Another advantage of this vector is conjugative transposon that can be excised from vector and inserted randomly in vivo in to another DNA molecule by purified Tn 5 transposase.

Construction of a conjugative, site specific integrative plasmid- transposon for horizontal gene transfer between *E. coli* and *Streptomyces* strains.

PPL1 plasmid was used to clone an integrase gene from phage φC31, resulting a plasmid pPAOI6 (Figure 3a, 3b). We tested several clones for horizontal transfer between *E. coli* S17.1 strain and *Streptomyces lividans* TK24 strain. The best transfer was obtained for plasmid pPAOI6 were orientation of integrase gene is in opposite orientation to the gene for resistance to apramycin. Conjugative transfer of pPAOI6 gene in to *S. lividans*

strain is confirmed by resistance to apramycine or G418. Additional confirmation of transfer was obtained using PCR method. We were able to amplify 2 kb insert using specific primers for the ϕ C31 integrase gene, and no PCR amplification was obtained for control *S. lividans* TK24 strain.

- 5 The goal of this construction was to produce transposon that is further used in functional analysis of the metagenomic DNA library from the soil that was constructed in a laboratory (Figures 4a, 4b, 4c).

Functional analysis of the metagenomic DNA library from soil.

The fosmids library consists of 120 512 clones, containing ~ 40 kb inserts of soil DNA.

- 10 The library contains approximately 4.8 Gbps of the DNA cloned from soil. Ten percents of the library was screened by using a PCR approach for the presence of the genes involved in production of secondary metabolites (PKS). Using gene-module specific set of primers we were able to identified positive clones organized in microtiter plates (96 wells). Sequences (based on PCR products) obtained from fifteen randomly positive 15 clones indicate that the DNA library contains very little sequence redundancy limited to one and that the sequences were found to be new and very diverse in comparison to gene databases (data not shown).

Two fosmids DNA was prepared from two positive clones (FS3-124 and FS3-135) and analyzed by sequencing. DNA analysis *in silico* shows high G+C contents of 72% and 20 69% respectively of the cloned inserts and presence of cluster genes that could be involved in biosynthesis of secondary metabolites (Fig 5 a and b). No specific phenotype was observed for the two clones in *E coli*. We employed pPAOI6 transposon mutagenesis to produce conjugative mutants. Transposon mutants FS3-124::pPAOI6 and FS3-135::pPAOI6 were isolated using apramycine as selective antibiotic. Obtained 25 mutants were then tested on LB pates containing chloramphenicol. About 1% of the tested clones are chloramphenicol sensitive. These clones contain transposon inserted into locus encoding chloramphenicol resistance gene and not into cloned DNA insert. ApraR and ChloS transposon mutants are then used for horizontal gene transfer into

Streptomyces lividans TK24 strain. Fosmid DNA was prepared from mutants and transformed into *E. coli* S17.1 strain. Horizontal gene transfer between *E. coli* S17.1 and *Streptomyces lividans* was done due to inserted pPAOI6 transposon. Transconjugants of the *Streptomyces lividans* were tested by PCR to confirm gene transfer and integration of 5 the conjugative fosmid in to *S. lividans* chromosome. Both transconjugants showed an increase in doubling time, morphological modifications and pigments production in comparison to the control (Figure 6).

References

- Brau, B., Pliz, U., and Pipersberg, W. 1984. Genes for gentamicin-(3)-N-acetyltransferases III and IV: I. Nucleotide sequence of the AAC(3)-IV gene and
5 possible involvement of an IS 140 element in its expression. Mol Gen Genet. 193 (1),
179-187.
- Guiney, D.G., Yakobson, E., (1983). Location and nucleotide sequence of the transfer origin of the broad host range plasmid RK2. Proc Natl Acad Sci USA, 80: 3595-8.
10 T. Kaeberlein, K. Lewis, Epstein S. S. 2002. Isolating "Uncultivable" Microorganisms in Pure Culture in a Simulated Natural Environment. Science 296, 1127-1129.
- Mazodier, P., Thompson, C., 1989. Intergeneric conjugation between *E. coli* and
15 *Streptomyces* species. J. Bacteriol. 171, 3583-3585.
- Pace, N.R. (1997). A molecular view of microbial diversity and the biosphere. Science 276: 734-740.
- Pridham, T.G., Anderson P., Foley, C., Lindenfelser, L.A., Hesseltine, C.W., and
20 Benedict R.C. (1957). A selection of media for maintenance and taxonomic study of *streptomyces*. Antibiotics Annual 1956-1957, 947-953.
- Voeykova T, Emelyanova L, Tabakov V, Mkrtumyan N. 1998. Transfer of plasmid pTO1 from *Escherichia coli* to various representatives of the order *Actinomycetales* by
25 intergeneric conjugation. FEMS Microbiol Lett. 162(1):47-52.
- Sambrook, J., Fritsch E.F. and Maniatis T. 1989. Molecular cloning: a laboratory manual, 2nd edition. Cold Spring Harbor Laboratory, Cold Spring Harbor NY.

Torsvik, V., Goksoyr, J. Daae, F. L. 1990. High diversity in DNA of soil bacteria. *Appl. Environ Microbiol.* 56, 782-787.

5 Torsvik, V., Sorheim, R. and Goksoyr, J. Total bacterial diversity in soil and sediment communities a review. *J. Industr. Microbiol.* 17, 170-178.

Torsvik, V., Daae, F.L., Sandaa, R.A., Overas, L. 1998. Novel techniques for analysing microbial diversity in natural and perturbed environments. *J. Biotechnol.* 64. 53-62.

10

Zaehner, H. and Fiedler, F. P. 1995. Fifty years of antimicrobials: past perspectives and future trends. In *The need for New antibiotics: Possible Ways Forward. Fifty -Third Symposium of the Society for General Microbiology*, Bath. England, UK, (Hunter, P.A., Darby, G.K. & Russell, N.J., eds), pp. 67-84.

15

Zechner, E.L., de la Cruz, F., Eisenbrandt, R., Grahn, A.M., Koraimann, G., Lanka, E., Muth, G., Pansegrau, W., Thomas, C.M., Wilkins, B.M. and Zatyka, M. (2000) Conjugative-DNA transfer process. In Thomas, C.M. (ed.) *The horizontal gene pool*. Harwood Academic Publishers.

20

CLAIMS

1. A method of analysing a library of polynucleotides, said polynucleotides being contained in cloning vectors having a particular host range, the method comprising (i) selecting cloning vectors in the library which contain a polynucleotide having a particular characteristic, (ii) modifying said selected cloning vectors to allow a transfer of said vectors and/or expression of the polynucleotide which they contain into a selected host cell, and (iii) analysing the polynucleotides contained in said modified vectors upon transfer of said modified vectors into said selected host cell.
5
- 10 2. The method of claim 1, wherein the library comprises a plurality of unknown polynucleotides.
- 15 3. The method of claim 1 or 2, wherein the library comprises a plurality of environmental DNA fragments.
4. The method of any one of the preceding claims, wherein the cloning vectors of the library are *E. coli* cloning vectors, preferably cosmids, P1 or BAC.
- 20 5. The method of any one of the preceding claims, wherein the selected vectors are modified by targeted insertion, into the vector, of a target polynucleotide construct.
6. The method of claim 5, wherein the targeted insertion is performed in a region of the vector distinct from the polynucleotide.
- 25 7. The method of claim 5 or 6, wherein the target polynucleotide construct comprises an origin of transfer functional in the selected host cell.

8. The method of claim 7, wherein the origin of transfer is functional in *Streptomyces* host cells.
9. The method of claim 8, wherein the origin of transfer is selected from RP4, pTiC58,
5 F, RSF1010, ColE1 and R6K(α).
10. The method of any one of claims 5 to 9, wherein the target polynucleotide construct comprises an integrase functional in the selected host cell.
- 10 11. The method of claim 10, wherein the integrase is selected from φC31.
12. The method of any one of claims 5 to 11, wherein the target polynucleotide construct comprises a transcriptional promoter functional in the selected host cell.
- 15 13. The method of any one of claims 5 to 12, wherein the target polynucleotide construct is contained in a transposable nucleic acid construct.
14. The method of claim 13, wherein the transposable nucleic acid comprises, flanked by two inverted repeats, the target polynucleotide construct and a marker gene.
- 20 15. The method of any one of the preceding claims, wherein the cloning vector comprises a first marker gene and wherein, in step ii), the cloning vector is modified by:
 - . contacting in vitro, in the presence of a transposase, the selected cloning vectors with a transposon comprising, flanked by two inverted repeats, the target polynucleotide construct and a second marker gene distinct from the first marker gene, and
 - . selecting the cloning vectors which have acquired the second marker gene and which have lost the first marker gene.

16. The method of any one of the preceding claims, wherein, in step (i), the cloning vectors are selected by molecular screening.
17. The method of any one of the preceding claims, wherein, in step (iii), the modified cloning vectors are transferred into the selected host cell by conjugative transfer.
18. The method of any one of the preceding claims, wherein, in step (iii), polynucleotides are analysed by determining the phenotype or properties of the host cell upon transfer or expression of the modified vector.
19. A method for the identification or cloning of polynucleotides encoding a selected phenotype, the method comprising (i) cloning environmental DNA fragments into *E.coli* cloning vectors to produce a metagenomic library, (ii) identifying or selecting cloning vectors in said library which contain DNA fragments having a particular characteristic of interest, (iii) modifying the identified or selected cloning vectors into shuttle or expression vectors for transfer or expression in a selected host cell, (iv) transferring or expressing the modified cloning vectors into said selected host cell and (v) identifying or cloning the DNA fragments contained in said modified cloning vectors which encode said selected phenotype in said selected host cell.
20. A transposable nucleic acid construct, wherein said construct comprises an origin of transfer flanked by two inverted repeats.
21. A library of polynucleotides, wherein said library comprises a plurality of environmental DNA fragments cloned into cloning vectors, wherein said environmental DNA fragments contain a common molecular characteristic and wherein said cloning vectors are *E. coli* cloning vectors comprising a target polynucleotide construct allowing transfer or expression of the environmental DNA into a selected host cell distinct from *E. coli*.

22. A polynucleotide sequence comprising all or part of SEQ ID NOs: 1 or 2, or of their complementary strand.
- 5 23. An oligonucleotide comprising SEQ ID NO: 3 or 4.

ABSTRACT

The present invention relates to methods for the identification or cloning of polynucleotides encoding a selected phenotype, particularly from environmental DNA. In a specific embodiment, the method comprises (i) cloning environmental DNA fragments into *E.coli* cloning vectors to produce a metagenomic library, (ii) identifying or selecting cloning vectors in said library which contain DNA fragments having a particular characteristic of interest, (iii) modifying the identified or selected cloning vectors into shuttle or expression vectors for transfer or expression in a selected host cell, (iv) transferring or expressing the modified cloning vectors into said selected host cell and (v) identifying or cloning the DNA fragments contained in said modified cloning vectors which encode said selected phenotype in said selected host cell.

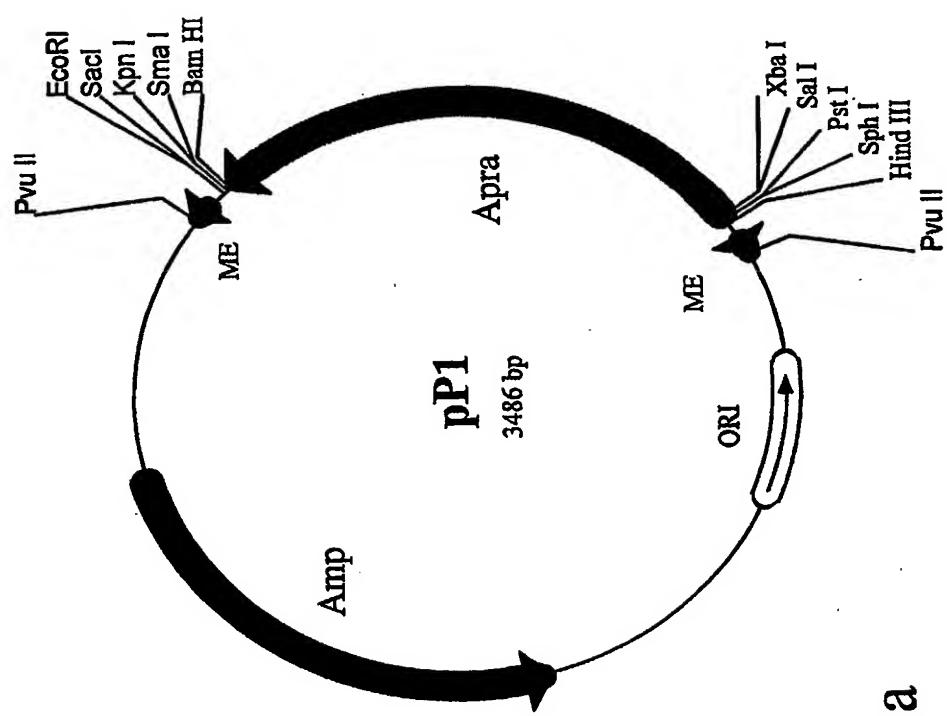


Figure 1a

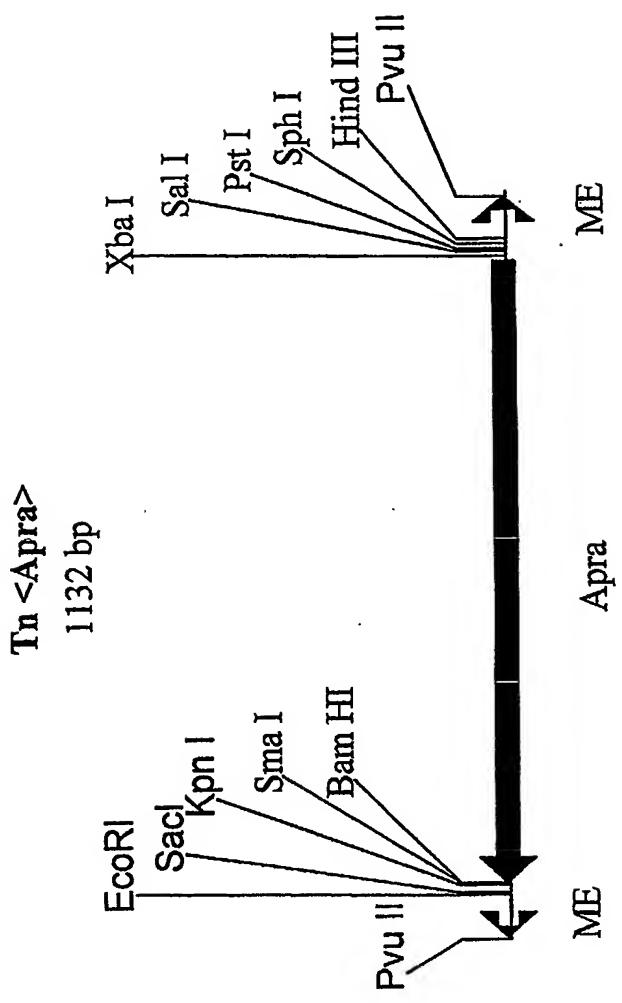


Figure 1b

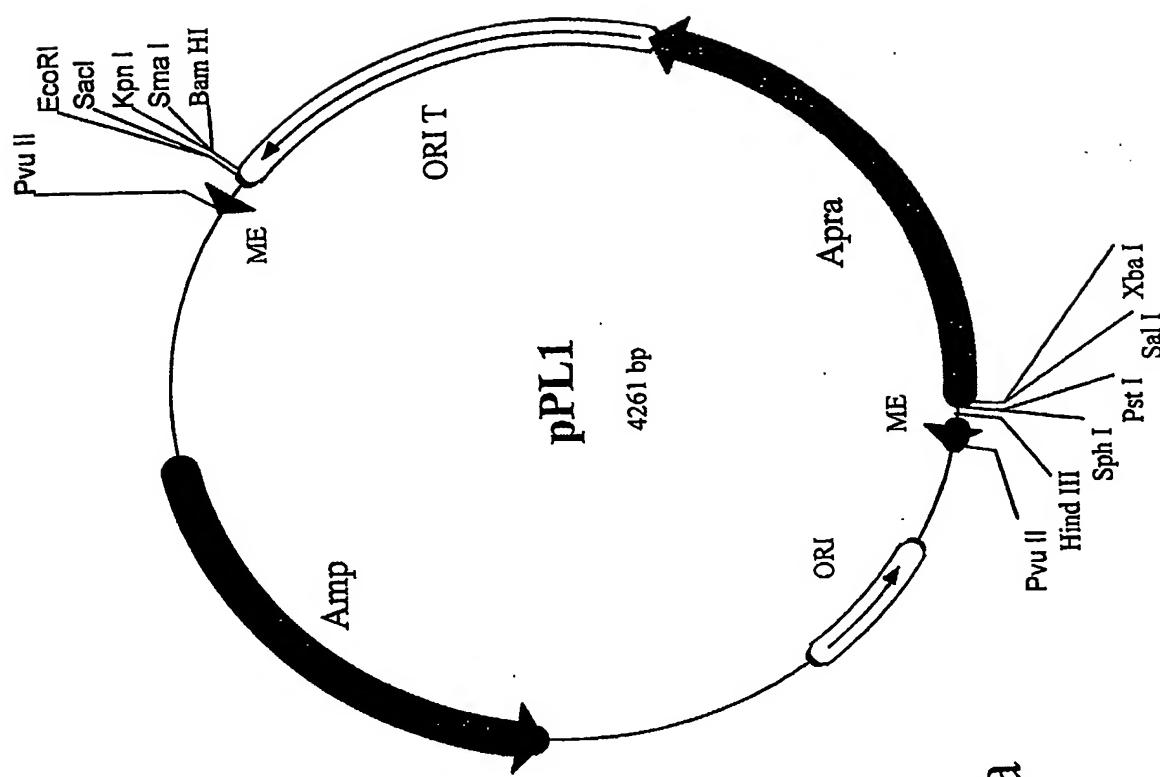


Figure 2a

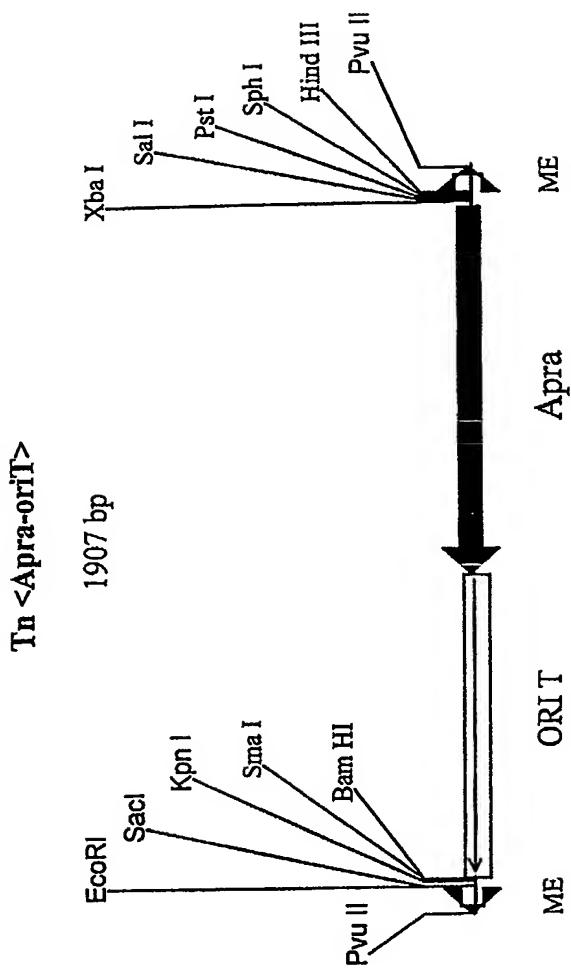


Figure 2b

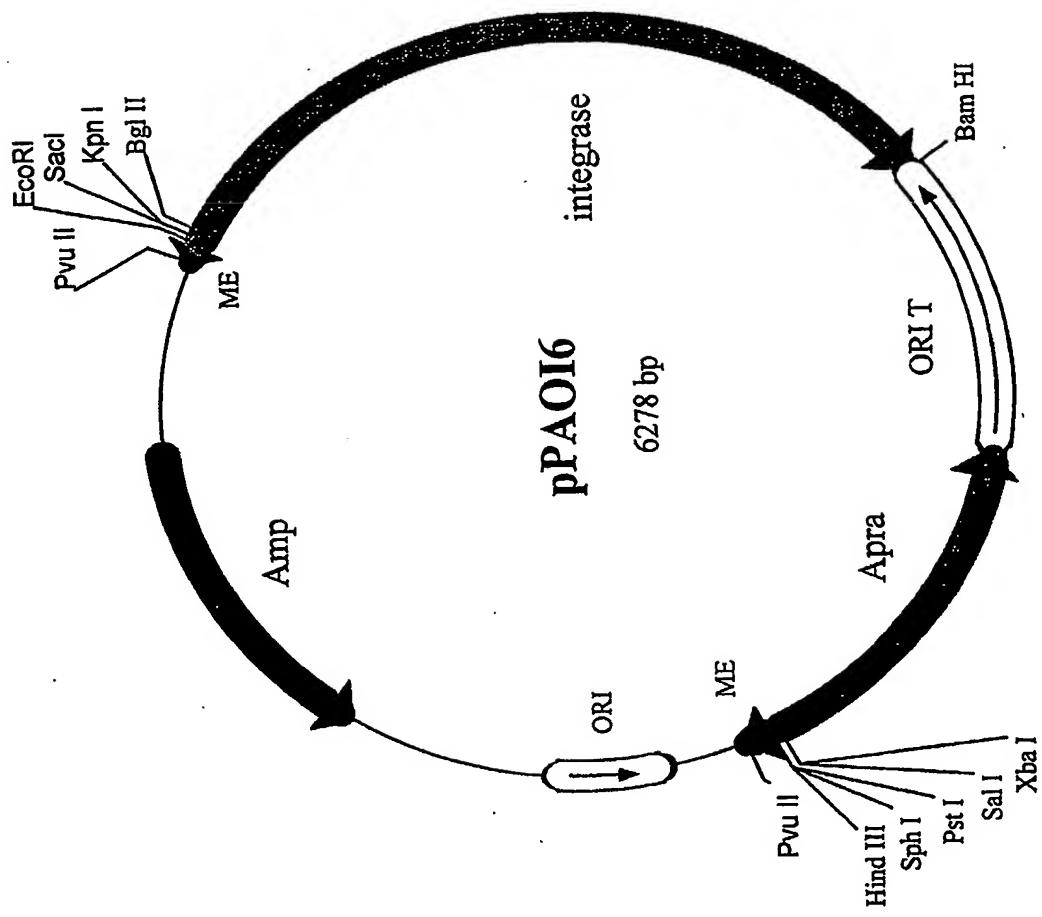


Figure 3a

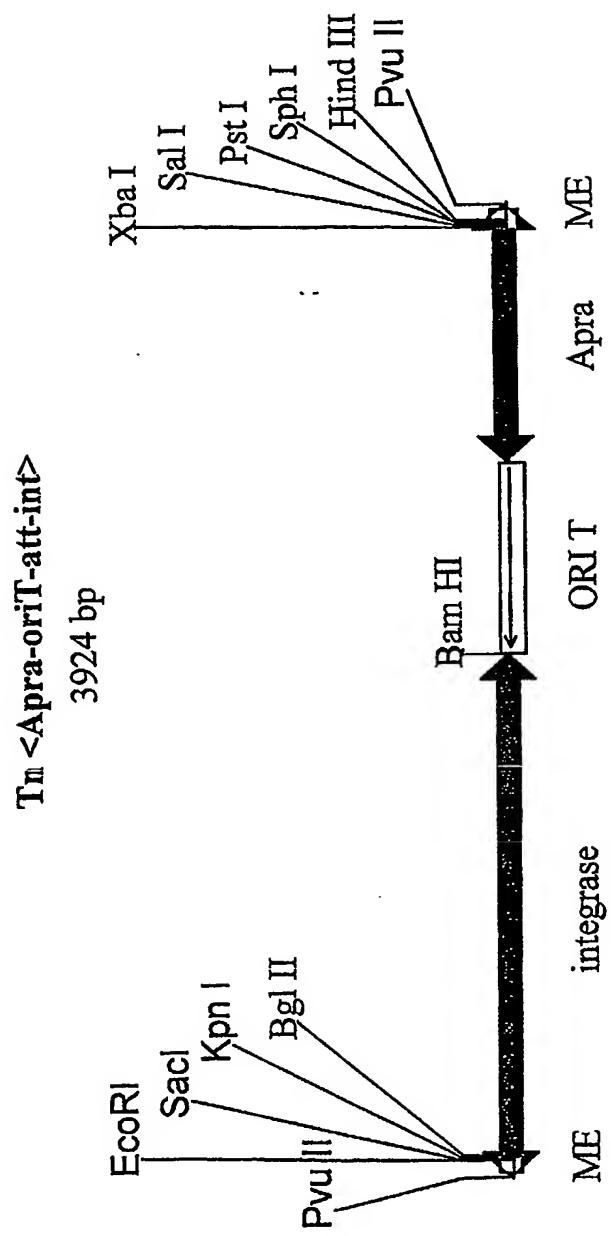


Figure 3b

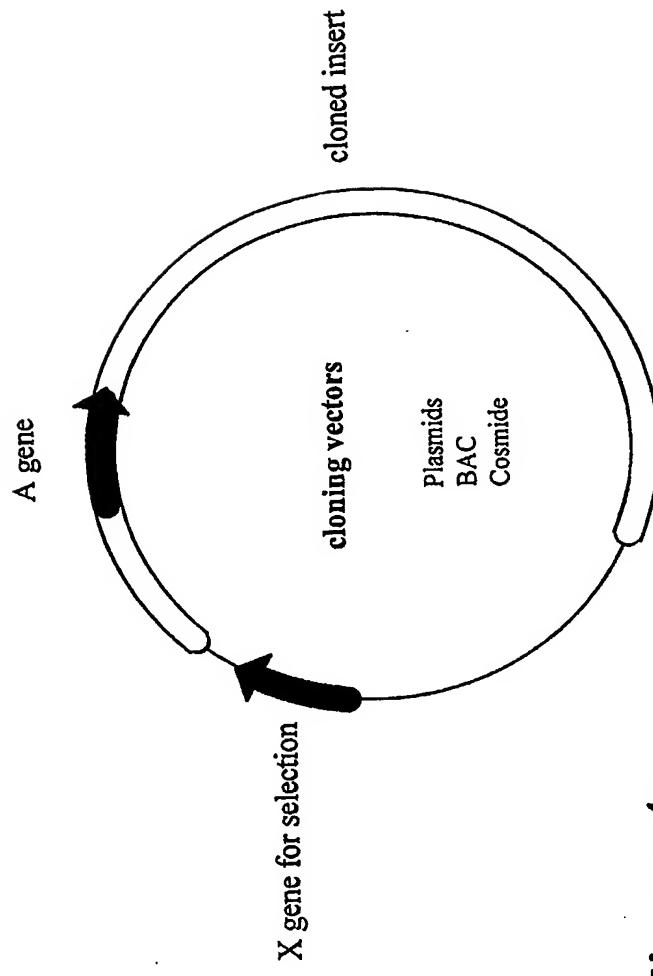


Figure 4a

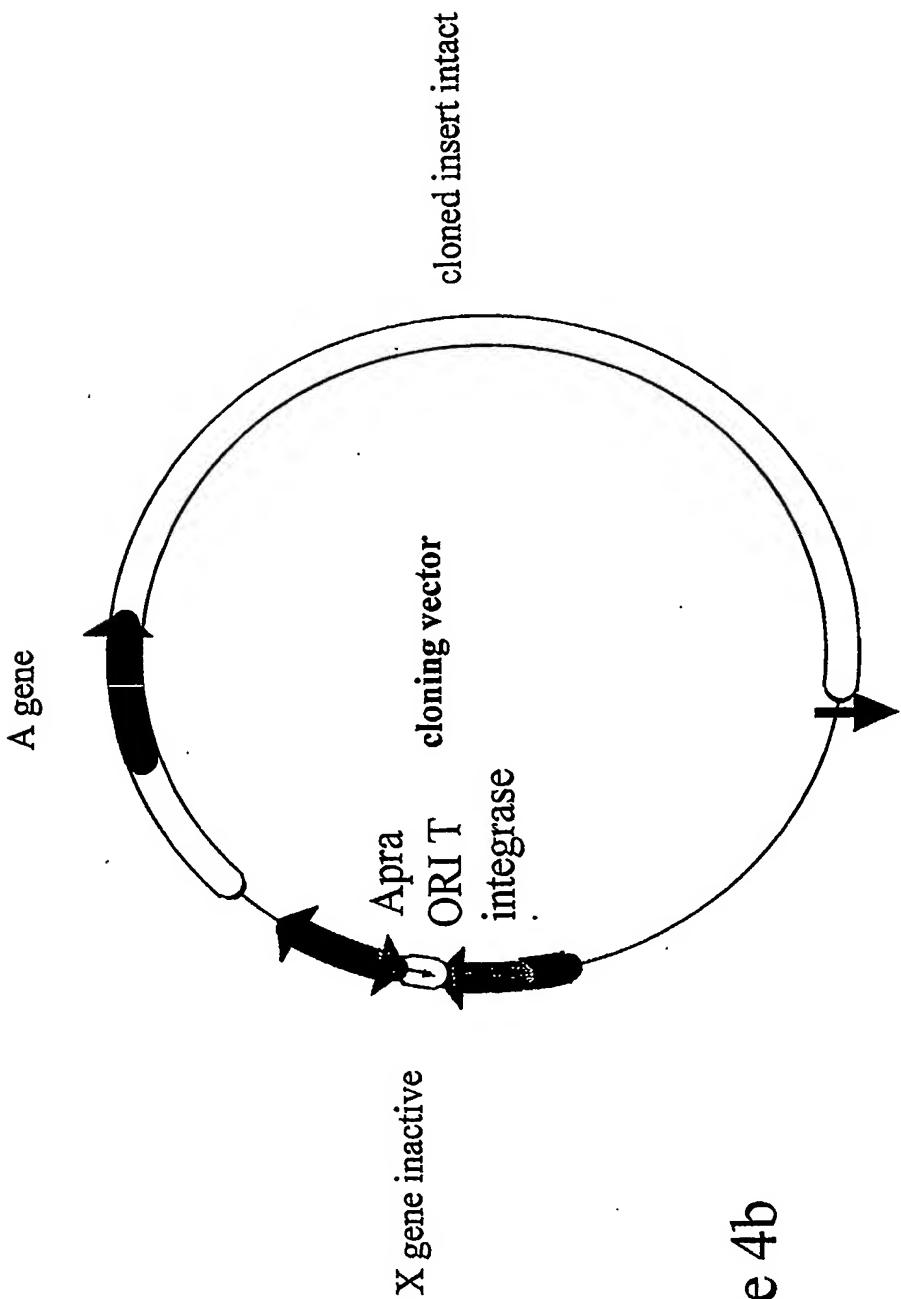


Figure 4b

Conjugative and integrative plasmid

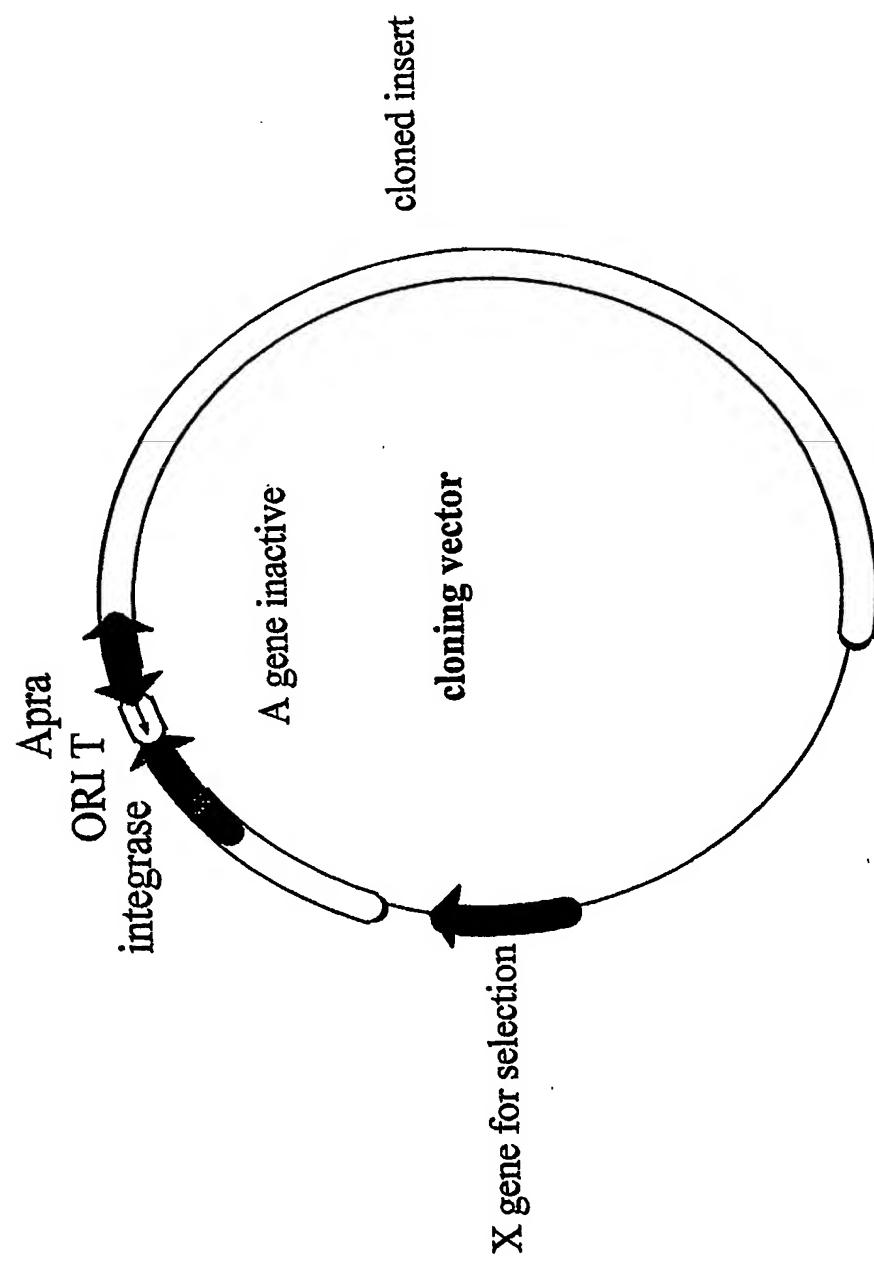


Figure 4c

Figure 5a

10/35

CDS 76..1134
/note="ABC_transportr"
/gene="TAP2 PROTEIN"
/blastp_match="Oryzias latipes"
/blast_score= 0.002
complement(1096..2430)
/note="none"
/blastp_match="Anabaena sp"
/gene="ALR1117 PROTEIN"
/blast_score=2e-18
CDS 1178..1624
/note="Gram_pos_anchor"
/gene="CELL WALL SURFACE ANCHOR"
/blast_score=1e-04
/blastp_match="Streptococcus pneumoniae"
complement(2506..3567)
/note="CONSERVED"
/gene="HYPOTHETICAL PROTEIN"
/blast_score=0.019
/blastp_match="Deinococcus radiodurans"
complement(2906..4222)
/note="glycosyl transferase"
/gene="lipopolysaccharide"
/blast_score=2e-23
CDS complement(4092..5321)
/note="glycosyl transferase"
/gene="glycosyl transferase"
/blast_score=1e-15
CDS complement(6337..8502)
/note="PUTATIVE"
/gene="GLUTAMINE AMIDOTRANSFERASE"
/blastp_match="Bordetella bronchiseptica"
/blast_score=1e-16
CDS complement(8181..9530)
/note="none"
/gene="MEMBRANE PROTEIN"
/blast_score=0.035
CDS complement(9531..10721)
/note="NOEC Transmembrane"
/gene="NODULATION PROTEIN"
/blastp_match="Azorhizobium caulinodans"
/blast_score=3e-07
CDS complement(10504..11274)
/note="PUTATIVE"
/gene="HYDROLASE"
/blastp_match="Streptomyces coelicolor"
/blast_score=4e-14
CDS 12874..13689
/note="HYPOTHETICAL Meth-transf"
/gene="PROTEIN PA1088"
/blast_score=2e-06
/blastp_match="Pseudomonas aeruginosa"
CDS 14195..15976
/note="PUTATIVE, Glyco_transf"
/note=""
/gene="LIPOPOLYSACCHARIDE BIOSYNTHESIS"
/blast_score=2e-06
/blastp_match="Vibrio cholerae"
CDS 15427..16512
/note="PATHWAY: INNER CORE LIPOPOLYSACCHARIDE
BIOSYNTHESIS"
/gene="PHOSPHOHEPTOSE ISOMERASE "
/blast_score=3e-17
/blastp_match="Helicobacter pylori"
CDS 15579..16253

CDS /note="none"
 /gene="PHOSPHOHEPTOSE ISOMERASE"
 /blast_score=2e-22
 /blastp_match="Neisseria meningitidis"
 complement (16505..17656)
 CDS /note="BIOSYNTHESIS PUTATIVE"
 /gene="LIPOPOLYSACCHARIDE"
 /blast_score=2e-17
 /blastp_match="Thermotoga maritima"
 /pfam_match="Glycos_transf_1"
 complement (17657..18697)
 CDS /note="none"
 /gene="ALR3073 PROTEIN"
 /blast_score=6e-27
 /blastp_match="Anabaena sp"
 complement (18615..19304)
 CDS /note="none"
 /gene="ALR4487 PROTEIN"
 /blast_score=8e-07
 /blastp_match="Anabaena sp"
 complement (19301..20596)
 CDS /note="ATP_GTP_A"
 /gene="ABC TRANSPORTER"
 /blast_score=3e-61
 /blastp_match="Synechocystis sp"
 complement (20535..21476)
 CDS /note="PERMEASE COMPONENT"
 /gene="POLYSACCHARIDE ABC TRANSPORTER"
 /blast_score=6e-41
 /blastp_match="Clostridium acetobutylicum"
 complement (22025..22951)
 CDS /note="involved in the synthesis of a polysaccharide
 capsule ?"
 /gene="32.3 KDA PROTEIN"
 /blast_score=3e-17
 /blastp_match="Sphingomonas sp"
 23155..26523
 CDS /note="peptide syntase"
 /gene="mcyA, mcyB and mcyC"
 /blastp_match="Microcystis aeruginosa"
 /blast_score=0.0
 CDS 26409..34433
 /note="polyketide syntase et peptide syntase"
 /gene="mcyD, mcyE, mcyF and mcyG"
 /blastp_match="Microcystis aeruginosa"
 CDS 34418..37500
 /note="CYSTATIN"
 /gene="PEPTIDE SYNTHETASE"
 /blast_score=0.0
 /blastp_match="Anabaena sp"
 CDS 35359..37500
 /note="gene cluster"
 /gene="nostopeptolide biosynthetic"
 /blast_score=2e-41
 /blastp_match="Nostoc sp"
 Sequence 37500 BP; 6199 A; 12698 C; 12769 G; 5834 T; 0 other;

gatcggtgcg	ggccttccg	ctattacgcc	agctggcgaa	aggggatgt	gctgcaaggc	60
gattaagttg	ggtaacgcca	gggtttccc	agtacgacg	ttgtaaaacg	acggccagtg	120
aattgtata	cgactca	tagggcgaat	tgcgactcg	tacccgggga	tcccacgtac	180
cacgagctca	tcttggaaag	cccgctccag	acacccgagg	aatggaaagcg	cggcgtcgtg	240
agaccgtaca	cgcagaagcg	tctcgatcg	ttcggtact	ggcttgcgac	gacgatgaag	300
ctggatctca	ccaggatgtt	cgccgcaggc	agctcgatgg	gccccatcggg	cgcgatcatg	360
ctcgcgattc	gctatccccgc	acgattcgcg	tggaccgtgt	cgtgggtcgg	cgtccacgtg	420
cccgccact	ctccgcaatt	cacatcg	tacgactgg	tgtacggccg	gccccacttg	480
aagggtccgt	tgcgagaacgg	cacggccggtc	tggatcatt	ttatgtacgt	ctggatcacgt	540
cggcagccatc	cgaggcagg	catcggttc	atcacgttct	cgaacggcaa	gaacgactcg	600
cgatcggt	ggggccaggc	cgtcgaaattc	ctgaagacgc	tgcaggagac	caggcaaccg	660
caccttgc	tctggggcca	ggagggacac	ggccagcg	cgaagatgcc	ggaagggtggc	720

atgcgcacgg	cgacgcccgtt	gttgtcacccg	agcgagtaga	cgtcgaa	gttcttgatg	9060	14/35
ccgggaatga	acaggagctg	accccagccc	tgcttggcg	cctcgaaccc	gaggagagc	9120	
gagatcaca	ggaagacgag	ccgcaaccgc	ttgacgtccg	tcgtgaggac	tgcgacagg	9180	
tacgtcata	cggtggactt	catgaagtgc	atcatgtgc	cccacgcgaa	gtccttggtag	9240	
ggcgcacgca	gcgtcgacga	caggctgtgg	agcccgaaaga	aggcgaggag	cgccagccgc	9300	
gcgtcgatcc	gcagcgcctg	cccgagacagg	aacgcgatcg	cgagcgtta	catgcccggcc	9360	
aggagcgaga	tgttcatgtg	gatgagaaaag	tcgttccaca	cccagagctc	cggccggaaag	9420	
tafcgcgtga	acaggttagaa	cagaatcgcg	tagaacggcc	cgcgacgcgc	gtgaaaggcg	9480	
ccgaaggcga	ggagggcggag	gacgaaagcg	gctcgaagca	ttgcgtgtg	ttacctgtta	9540	
tcggtccccg	tggtccagag	cggctcgctc	cgggtctctt	ctacattgaa	caccatgttag	9600	
agggccggaa	tgcaggtgaa	catgagcacc	acgaaaagaa	tcgcgcacac	caccatgttag	9660	
gcgaagaacc	cccgcgtctt	gtagagcttc	tccgattct	gcacggggct	gttccggcagc	9720	
atcccgatgt	gcaggttagta	ggcgaacatg	cccgcggca	cggcgcgaa	caggatcagc	9780	
tcgaggttat	agcggacgat	gaagatccgc	ccgaagagg	cgcagcagg	ggcgtagaac	9840	
accatgtca	cgagcagccg	cttcctgttg	tagaaggcaa	acgactttcg	gtacgaggcc	9900	
gacgaccgac	cgttgcgtat	gtgcccggaa	tccgcgtacc	gcttcgtcg	catgaagaac	9960	
gcccggccca	tccactacga	gatcacgagc	gacatcgccg	gcacgcgatc	ctgaatcagc	10020	
gggaaccagg	cgaggaggag	ccggacggcg	ttttcgcgg	actcgctcg	cacgtcgagg	10080	
tacggccact	cctcgtcg	aatggggcgc	acgttgcacg	tgacgcgcag	caccagagc	10140	
gccagggccg	acagcgcgaa	gtatctttc	acgacgcgcg	cgagcacaa	tccggccacg	10200	
cccacgagga	tccactcggt	gtagcccccc	gcccgttga	tcttccggaa	gggcacccggc	10260	
cggtgcgtt	tctccggatg	cagcagatct	ctcgggccgt	ccaggagctc	gttcaagcag	10320	
tagttgtcg	aggcgatgag	gcaggtcg	gchgagcgcga	gchgagcg	cgggacggcg	10380	
gtccagccga	agagctgcgg	ttcgttagaa	aacgcgcagca	gcacgcgcag	cagecatgaac	10440	
gcgttcttga	accagtgtatc	gatccgcgcg	atctggacgt	acggccaaat	cgcgaaaccg	10500	
gagctagcgg	ctgacgacat	cgttgcgtc	ctcgacgcgc	cggatcgca	accatccgcg	10560	
gccggctgtg	acggccgtgc	ggccggcgat	gaggcgccag	gagactctg	cggccatcgc	10620	
tccttcgtcg	tccacgtcg	gtcgcattttc	gacatacagg	acctcccgag	gttcccgatgt	10680	
ccacatctca	cacgcacat	ggaagccgcg	cggtgcggc	ttcaggcggt	tcacgttgc	10740	
ggcggctgtg	cagagcgcca	cgaaaaaaatg	ctgcgacgtg	ccgagcgcgg	cgagcttgc	10800	
ctcggccgcg	tagtccgaca	gcacgcccgg	cctgagcccc	gcccgcgaa	gchgactcgag	10860	
cgcgcccggc	aacccggggc	gccccccacca	ccgcagatac	ttgagcgcc	ggcgcacccat	10920	
ccattcggtt	accgtcgccg	ccacgcaccc	gttgcggcgc	ccgagcgct	gcccgggtcg	10980	
cgcgatctgg	cgcgcccgcg	gccccgtcg	ggccggcgccc	aggcgcggaa	gatcctcatg	11040	
cgcggccgg	tactccgc	cgatgcgcgc	cgtgtccact	cgcggccccc	acccggacgt	11100	
cacgagcgcc	gtggcggcga	gttccgcgtc	catggcgccg	cgcaacggac	cctyccggta	11160	
cagcgtccg	tccacgtcg	acagcaegg	cttgaccccg	cgcaacgg	tcactctgcg	11220	
aggcccaagc	tccgcgaaat	cgccgcaccc	tttgacgcgt	cgcgaaggaa	cacgtcacga	11280	
accacacagg	cagccagcc	cgccaaacagg	accggccgtg	ccgcgcggcc	tgccgtccgc	11340	
cagctccgc	gcacggccgc	cgtgtcgatc	gccccggcgc	cgtagcagag	cagaatcggg	11400	
acgagccgca	ggtgataccg	gggatgtccg	aagacgatcg	cgtgaatgc	gctcatgaac	11460	
acgtacggc	tcaggagcag	caggtgcgcg	cgccagttga	ccggccggcgc	gaacgagatg	11520	
ccgatgacgc	cgagcaccat	gaccgcgcg	tacgatccca	tgatggacac	cgccccggcg	11580	
aagaacagcc	accgcggccg	acggtagaac	ccattccga	cgccggcgat	gaagtcgcgc	11640	
tccagccccc	agaagtctgc	gaacttcagc	accgaacggc	gaagcgtcg	tcccggtatgg	11700	
gccttcatgt	agtctgagcgc	ctggcgctgc	gccccacttc	ctttcggtcc	ctcgttccag	11760	
cggctgcccgt	ccggcgcgcg	aggcggatcc	gtgcgcgacc	atcccttgc	gcccgtgagc	11820	
gagatcgcgt	cccacatgcg	atctccggc	gttacgtcg	atgtcccat	catcaggatc	11880	
agcccgccga	gcgtgtccac	gaccgtgaac	gagcgttgc	ggagcgtt	gcmcacgctc	11940	
cacggcccg	cgatcgccgc	gtatccggcg	aagagaagca	cgcccaacgc	cagacgcgcg	12000	
cgaagcggg	tgcggccgacc	gacgagcgcc	gcccgcgtga	ggatcgccac	gtacggccac	12060	
atgategac	gaacgagcgc	ggacgcgcgc	agcgcgacgc	cgctcgcgag	cgcgatccac	12120	
gccgagtgc	gcctgtatgg	gcgatcgatg	agcgtcagac	atcccaagag	caacaaaaga	12180	
aggagtgtg	tgaagagcgt	ctcgacacgc	accagcacgc	cggagaacag	cagcgaggga	12240	
tagaacgcga	acgcccggcc	ggcaatgagg	ccggcgcgtg	cgctgaacaa	cctgcgcggcc	12300	
atagagataca	cgagccagac	accgagcagg	ctcaacggaa	cctgcgcgg	ccggaccggcc	12360	
gtcagcgatt	ccgagcccg	gaccgcggag	acgcccgcga	cgaaacgcggg	aaagagcgcc	12420	
gccccggatgg	acgtcggttc	ccccggcccc	cacgcgaacc	cgttcccgcc	gacgacgttt	12480	
ctggcaagcg	tgcgttagt	ctgctcgctg	cgaaatctcg	gcmcgcacgt	cttgaaacccc	12540	
caaatacggc	cgagccggag	gaccaggggc	agcgtgagga	tgacgagaat	tgcccggcgc	12600	
tcccaatggg	tgcgtctggc	ctcgctcg	ccggggggcg	cgattggca	ggggggccac	12660	
gatgggagt	ggggagccgt	catcatcg	taaaaacaaga	agttacggc	aaacctgtgc	12720	
cacccggggaa	tcaagaaaaga	ctctgcacac	gggtccggcg	ggggcgcacgc	ccggcgcgcgc	12780	
ttttccggcc	ggaagtcggc	ccccaaaggcg	acccggctgg	gcacagccca	gtagaaatct	12840	
cattctggga	tgtattttgc	ccatgtatca	tgaccatttc	gaggggcctt	cgtgacgccc	12900	
tggggatggc	ataagcttgg	cacctgcgt	ggcgacgtgt	cgatgatcag	cgccgtgag	12960	
gagctccaga	agccctatcg	ggacgacccg	gtgcggccgt	aatacgtggc	gcggcggttc	13020	
cagtccaccgc	tggggccgcgt	cctccattcg	cgtcagatcg	gggtcgtcg	agagctcgtc	13080	
cgcgccgcagg	gcatccggcg	cgccggccggag	atccgcacccg	gtccagcg	gctgaccgtg	13140	

cggccacgcg cgccgcgcg gcctgcagcg cgggtctcat gaatccgtac acgacaggat 17340 16/35
 tgcacgtt gaccaggcgc cgccgcacga gacccaatcc cggaaagcatg cgccgcacacg 17400
 cggcgagctg ttgcaggcgc cggccgcga gcggttcag gtcatacatc cgccgcgg 17460
 cgcctccgc gtcgagcgcg gcgcggcag gaccctcgcc ttaatccat ccgtgcacg 17520
 ggacgcccgcg gctttggcg gccccgaca accggatcgc cagcaccg 17580
 ccagatgtga gacgaggatc tcgtcaggc gggggactg attcatggtg ggcgcgatg 17640
 cggccggcgc gccagctac ttactcgcg tcgcaggatcc cccgcgcgg tagagcg 17700
 gacagcaggct gcaagtccc gacgggagga agcacacggc gtaccgcgtg agatcg 17760
 gcccgcattcg caggcccgc tgaccgaacc gttccgcgc cgtgaaaac tccgcgcgt 17820
 cgtactcgag gtggccgagc gcgagcaggaa gacgatggc atgatcgatg gcccgcgt 17880
 gcgccactg atcgctcccc cggccaggaa accgctccag ctgcctgacc atgcctcg 17940
 gcatcacgt ggacgactt gacatgttg tgcgtcgc cctgtaatgc gtcagggtt 18000
 cggatgaa accggatggg ccctcgcc agaggcgcag ccagaggatcc cagtcctcg 18060
 cggcgtggcg gccggatggc aacatctcc gccggagcac gccggcttt accatgacc 18120
 acgaaggcggt cacggcgatcc cgcaggacga ggtcgaggag acagcgcccc tcgagcggcg 18180
 gaagatcgag cgtgcgcagg acctcgccgg cctcgtcgat gacggcgcgg cccgtatgaa 18240
 tgaagctcca gtcggcgtgc tgcggagaaga gatcgacccg cccgcgaatc ttggccgat 18300
 cccacgatgc gtcggcgtcg agcagegcac cgaattcccc tcgcgcgcg gcatggc 18360
 gattgcgggc gccggcaagg cccgggttgg cctgagatg caccgtatg gcccgcgcg 18420
 atctcgccag gatgtctgcc tgcgcgtcg tcgatccgc gtcgcacgcg atacttcg 18480
 cgggctgata ggtctggag agcgcggact gcacggccct ggcgtatgtg cgcgcggat 18540
 tgaacaccgg gacgatcacc gacacgaggc cggggatatac ggtctggcg gccggcgagg 18600
 gctctcgagc ggcttcagcg gccatggccc tetgcactgg cgtcgcgcg gAACAGCATG 18660
 acgtttctcg aaagccacgg agacaacggc ggcgttca tcgcctcccg cagcgtctgc 18720
 gatgcggggc cgaggatgtg gacccgcgt gccgagaaga tcttctgcca gtactcg 18780
 gctgttctg tgatgtggcc cattccgcgc tggccggct ggcggcgct gaagagcacc 18840
 agcggcgccg tgcgtcgatcc cgcggcgacg agttctcgca aacggcgccg cggaaagtgt 18900
 tcggcgaccc cgaggctcat ggagaggatcc gttgcgtccg cggccgtgga caccgcgtg 18960
 agatcgaaatg agttcagcg gacgcgcgcg tttccgcgg ctgcgtactcg cgcgtactcg 19020
 gatcgcttgt agccggccgc cttcacgcct cgtgcgcagg acgcgtgcac gaagtgacc 19080
 gtgcgcgcgc cgaagtctgt catcgatcgatcc tgcgcgttgc gcaacgcgcg gacccatca 19140
 acgagcacgg ggcgcgcgc ggcggcgacg tgcccgaaaca gatccatgc gcccgcgttgc 19200
 tacgtccatc gtcggagcc cagccgttc cgcgcggg acatctcg tccggcttgc 19260
 caacgcggcc accagtagag cgcgttgcgc tgggtcaca tcatgggttc gtcggcgccg 19320
 cgcgcgaaac ggcgcgcgc gcatgcgcgc acgcgcgcgaa ctccacatcg gccacgcgcg 19380
 cccacgatcc atgttcttgc accgcgcgcg tgcgcgcgc acgaggaaacg 19440
 tctgggtcgg gttgtaaaac acgtgcaccc acaggatgtg tggccgcgc acgagatcg 19500
 cgcgaaatg gaggtcgagc gagaacttcgc ccatgaacgc ggacgcgtcc cccgagatccc 19560
 gccgggttat gttgcgtcg tagacgacgc gctgatcggt ggagcggtgc accatgaaac 19620
 cgaacgtgac gtcgtccagc gtcgtcgatcc ttctcgccgt gacctggagg ttcagctgt 19680
 tccccggcgt cacttcgtg tcgtgcgcgc acgacttcgc gccgacgaaag acgcgcgcgg 19740
 acacgatcc gacggggcgc cgcacactcg gctgtccgc ttctcgac ggcggacgt 19800
 acgcgtcgag gacttcgtatcc ggcgtgcgg cggccctcgc cgcgcgcgc agatacaccg 19860
 catgtcgca gagggtcgcc accgcgtca ggttgcgc gacgaggacg atcgcgatgc 19920
 cctggcgctt gaaacttcgtc atgcgtatcg acgacttcgt ctggacgtc atgtcgccgc 19980
 cgctcaaac acgtcgatcc atcgtcgatcc atcgtcgatcc cggccgcgc gtcggccgtc 20040
 cgagccgcgc gttcatgccc gacgagtagc gttgcgcgg cgtgtccacg aagtcgtgt 20100
 cgcggcggaaa ctcgcacgtc tcgtcgaaacc gtcggcgat ctcgcacccgt ttcatcccc 20160
 tgatcgacc ctggaggaac acgttctcgcc gcccggatgg gtcggatgg aacccgcgg 20220
 cgacttcgtat gacggcgccg gcgcgcgc gacgatcgac gccgcgcgc gtcggccgtc 20280
 ggtatctcgat cgcacccgtt agggtggatcg acttgcgcgc cccgttgcgc ccatgtatgc 20340
 cgagccgcgc ggcgcggatgg acttcgtatcg agacgtttct gaccggccca aactctgtt 20400
 cgtccagcgcc gtcggcgccg cgcggatgg acacgcgcgc cgcgcgcgc ggcacgcgc 20460
 cccgcaggat gtcgtacgc tcggccgcgc ggaacttcgtt ggagacgttgc tcgaagacga 20520
 cggggccgtt cacttcgtatcc gtcgtccgc aatttcgtatcg cggacgtatcg gacccgcgc 20580
 cagccgaaatg cgacgtcgatcc gacggatggc acgttcgtatcc cgcgcgcgc gacccgcgc 20640
 ggctgggttca ggcgcggatgg acttcgtatcg acgttcgtatcc cgcgcgcgc gacccgcgc 20700
 tgcggccgcgc actggatgttgc ggcgcgcgc ggcgcgcgc gacccgcgc gacccgcgc 20760
 aacatccaga cggcgatcgac gacccgtatcg aggtatccat cgtcgccgtat gacccgcgc 20820
 gccatcgaca ggcgcgcgc gacccgtatcg gtcgtatccat cgcgcgcgc gacccgcgc 20880
 ggcgcgcgc ggcgcgcgc gacccgtatcg acgttcgtatcg acgttcgtatccat cgcgcgcgc 20940
 accacggatcc cgcgcgcgc gtcgtatccat cgcgcgcgc gacccgtatcg acgttcgtatccat 21000
 tccctggaa agtacacccatcg cgcgcgcgc gtcgtatccat cgcgcgcgc gacccgtatcg 21060
 aagcggagcg cgcgcgcgc gacccgtatcg acgttcgtatccat cgcgcgcgc gacccgtatcg 21120
 tacgcgcgcgc gtcgtatccat cgcgcgcgc gacccgtatcg acgttcgtatccat cgcgcgcgc 21180
 ttgtatcgatcc gtcgtatccat cgcgcgcgc gacccgtatcg acgttcgtatccat cgcgcgcgc 21240
 acgcggatcgatcc gtcgtatccat cgcgcgcgc gacccgtatcg acgttcgtatccat cgcgcgcgc 21300
 atctcgcgcgc gtcgtatccat cgcgcgcgc gacccgtatcg acgttcgtatccat cgcgcgcgc 21360
 cccggccggat cgcgcgcgc gtcgtatccat cgcgcgcgc gacccgtatcg acgttcgtatccat 21420

ggtcgccgccc cgctcggcgg cgacgtcgcc gtttcgctc gacggcacgg ggctcgtaac 21480 17/35
 ggggacgagc gcggcgtccg atcggtttag ccgcacatcgcc tcgtacacca tgtagatctt 21540
 gctccagtagc tgcgtggcga tcacgcggcc gtggctctgg taccagtata gcaacgatgg 21600
 cccgtcgatc ccccttctggg tgacgtatcg gttgatctcg atgggctcg tgcggccccgg 21660
 cacctttagcc aggcgcggc cggtgatata gggctccca gggcgccccgg gcaggcagtt 21720
 catcgccgag tggatctgt cgccttcgg ctggctctgg tagtagccg tgcgtggc 21780
 cacgtaggcc tgcgcggcc ccaggtatag cgggttgatg tactcgcca cgccgagcac 21840
 cgccatgacg tccttggta agggcgcgt gtcgcgtccg gtccaccgct ccattggaa 21900
 cgggacgttc gcgagcggt aaacgcgcgc gtggcgagga aacacacgt gagaatgatc 21960
 ggactcacac cgtggccggc cccagaccac gacaacaacat gacgtcgcc 22020
 tggaaagaa cccctgcgcc ccacacgcgc gcccgtccg gtcgtggatcg tcatgaatct 22080
 gcatgcgcac gcgoggatcc atgacgaacg cggcaagaaa 22140
 tgagcgaccg gatccctggc cggcctcgg ccacttcgg cggccacacg 22200
 tcacgttgc ttgcgcagc accggatgc tgagcgcggg 22260
 gagacggccag gagtcgcagc gggaaatgcga tctgattgaa 22320
 tcatcagcac caggaacgcg agcgaaaaacg cgagcaccca 22380
 gaaacacgat cgcccggtt agcgaggcta gtatcgacac 22440
 cgccgaccaa ccccgcgacg agcagcgcga gcccggcagag 22500
 cgctcgccgt gtccggcagc gccgagagac gcgcggccggc 22560
 cgagcggtac gatcagaatc ccgtgggagt agtgtcgcc 22620
 accgcacgat cacctgccc taaaggagca gggaaactcac 22680
 ccagtcggc ctgaagcaca ctcatagcga tcgggacgaa 22740
 ggcagaactt gggcgccgaa acacgtaaacc atcaacgtca 22800
 gacgggattt gccggctggg gggggctcgtaa gatctaccc 22860
 atacaactct cttttact gcttagcaca gaaatgtct 22920
 aaatgtcatc tgcaggaaag agtaatgtct ctggccacag 22980
 ccgacagaaat cgctgtaccac cgctgacttgcg atccggcgag 23040
 gaaaactgca ttacaggaaac gccaggtcgg cagccggttc 23100
 ggcattgcag ggcgtttcggt ctgcccgggg ggggaaccgt 23160
 tggagcaggg cagggacagg catggatgg tccgtgagac 23220
 cgggctcgcc gggagctcgtaa gggccgtccg 23280
 gcccggccg taagtttga cggcgtctcg ctctcgatc 23340
 aacgcgetcg cgacccgcct gcaagcgtcat ggcgtcgatc 23400
 tccctcgacc ggtccatcga gtcatcgta gggatctcg 23460
 gcttaegegcg cgctcgatcc tgcgtatccg aaggaccgco 23520
 gcccggccg ggatecgatc tctcgccgta cggcgtgagcg 23580
 gtcgagactca tctcgccatcg tgcggacgag cccggggcg 23640
 cccgacccgcg gctcgatct cgaccacctc gctacgtca 23700
 gggccggccg aggccgtcg gatgccgtatcg cggccgttgg 23760
 gtcgacgcat cgcccgatcg gggccgtccg cggacgttcc 23820
 gacgtctcgat tcaggaaat ctttgcacg ctcgcgaccg 23880
 tccgaggaca cgccgggggatcgatct cgaccacctc gctacgtca 23940
 gcccggctgt atctcccgatcgatct cggcgttgg cggcgtccg 24000
 cccgatctgc cgccgttgcg tgaggtatcg accggccggcg 24060
 ggcgtctcc agtggatcg gggccatcccg cagtgacacg 24120
 gcccggccg atgtcgatc tgcgcacacg ctgacttagcc 24180
 ctgcggccga ttggcgatcgatct cggccgttgg tgatcgatc 24240
 cagcccgatcg cgcacccgcg cgtggggatcg atctcattg 24300
 ggctacttgc atcagccgca ggcacccgcg gaaacggttt 24360
 gatcgatcgatcg tgcgtatcgatcc ctcgcatcc 24420
 ttcgtcgccg ggacggacgatcgatct ccaggtaag 24480
 atcgagaccg tcctcattcgatcgatct gtcggggatcg 24540
 gacgcgtccg gcaagcggatcgatct gtcggatcgatcc 24600
 accggagccg aggctcgcc acaggatcgatcc ggtatggcaga 24660
 cggccggccg cggactcgac ggtatcccgatcgatct 24720
 acggggatgc cctacaccgcg ggcggagatcgatct 24780
 atccgtcgatcg atgcgcacccgcg cggatcgatct 24840
 ttccgtctcg cgccccactcg cgtgtatcgatct 24900
 gaccacgtcc gggccacgcg gacggccatcgatct 24960
 gaggccggccg actttaccgg cttccggatcgatct 25020
 gtcgcgcgcg atgtcgatcgatct cggggactatcgatct 25080
 accgtcgatcg cggccggatcgatct catctcgatcgatct 25140
 atgttcacg cgtcgatcgatct gtcacgaaatcgatct 25200
 cgcgacgcgcg tgcacgatcgatct gatggcgatcgatct 25260
 ttgcacgccc tcatggacccgcg atctcgatcgatct 25320
 ggtgcctcg acacggatcgatct gaaacggatcgatct 25380
 ggtgcctcg acacggatcgatct aggtacgacgatcgatct 25440
 ggtgcctcg acacggatcgatct tgcgtgtccac ggtggccggatcgatct 25500
 ggtgcctcg acacggatcgatct aggtacgacgatcgatct 25560

25620 18/35

cgccggggccg ggcggcgtcc gtacgtcgcg ctcgactggc gtacggagcc ggtcggcag
cgtcagctcg ggcaactgct cagcgcccgat ggcgggacg cgatcgagat caggaacgtg
accaacgcgc ggctcggtca gcccggcccg tgcttcgatc tcgtcaagcg tgccgactgg
accggcacgg cggcgagac gcttgagagg gcccggcg acacgcgcga cgcgatcgac
ccggaatctgt ggccggctct cgcegaacgccc gccggctacg acggcgagtt cacctggtc
gcccggatcg gcccggatcg cttecgatcg cggttcgagc gcccggcact cgcgtcgctc
gcccgtcccc tgactcgatc cggcggtccc gcccgggagc cgccggccat tcagcgtac
gcccggatcgatc cgcgtcccgta tggcggtggcg cggcgatcgatc cgcggcgatc
ctcgaggagc gcgtgcggta tcaacctggt cggcggtct tcgtcgatc cgcgtcgatg
ccgctcacgc caagcgggaa gatcgatcg cggcgcccttc cggccggccga gagccgcgt
cccgatctgg acgtggcggt tgccaaagccc ggcacggagc tcgaacgcga gatcgccgc
gtctggcaga cgacgtccaa gatctcgatc gtgggattgc acgacaactt ttgcgatctg
ggccggcact cgcgtctgt cgcgcaggcg ttcgagcgta ttccggcgct cgtccggggc
aagcagtgtt cgatgtatcgatc gatgttccag taccggacgc tctatcgatc cgcgggttt
ctcgtgagc ggcggacgc cggcgccggg cagcttgcg gcgcgcaggaa tcgcggacgc
cgccaggcgc agatgtcgac ggcgtcgagg cctgcctga ccaaggtcgaa aggggtccgg
taatcgagaa cggcatcgcc ctcgtcggtt tggccggccg cttccggggc gcccgggtcg
tcgaggagta ctggcgcaac ctcgtcgccg gacacggatc gatcacggtc ttccacaaacg
aggagctggc gacggccggg atcagccaga gcgatctgca gaatcccgac tacgtcccg
cgccggggct cgtccggat cccgacatctc tgcacggcgc gttctcgcc atcagccgc
gccaagccga gctgtatggat ccgcaggaccc ggtcttcctc cggaggcgatc tggctcgcc
tcgaacacgc ggcgtatcgatc cccggcgatcgt tcccggtct categgtcg tggggggca
tgagcacggg gatgacgaaac acgacgtacc tgctgtcgaa ttccacacg caccgggat
tgctccagcc ggaggacgtc ctgcggcgc tgctcgcaaa cgagaacgcac tacctcaca
ggccgtctc gtacaatcgatc cacctcgatcg gaccggatc gaaacgtcgatc acccggtct
cgacgtcgatc ctcgtcggtt gtcaggatcgatc gtcagtcgt gtcacatcgatc cgtgtatcg
ccgcgtctcg cggcggtcg tccgtgtcgatc tcccgcaaca ggaaggatc gtctacgtcg
agggcgccat cgggtcgcca gacgggcact gcccgcgtt cgatcgac gcccaggca
ccgttccatcg caacggcgtc ggcatacgatc ttctcgcaag gtcggaggac ggcgtcccg
acgggcagac gatctacgcg gtatcccgatc gatcgccgc gatcgac ggcgtcccg
aggtcagttt cgcaggcccg agcgtggacg gccaggcgatc ggtcgtggcc atggcgaca
gcccggccga cgtggcgccg cagacgtatcg actacgtcgaa ggcgcacggc accggcacgc
cgctcgagatc tccgatcgatc gtcaggatcgatc tcccgatcgatc gtcacatcgatc
cgccggatcgatc tccggatcgatc tcaagagacaa ttccggggcact ctcgactcg
ccgcgtcgatc tccgtgtcgatc tcccgatcgatc tcccgatcgatc ctcgactcg
cgacgtcgatc tccgtgtcgatc tcccgatcgatc tcccgatcgatc tcccgatcgatc
tgaacgcgc gctgcgatc tggccgcgc tgccgcattcc ggcacgcgc gccgtgagct
cggttccggat cggcgccggacc aacgcgcac gtcgtgtcgatc ggaaggcgccg cccggccgg
agcgcacccgg cgcggccgc gccaggcgatc tgctcgatc tccggcgatc acggccgg
cactcgacgc ggcggatcgatc aacctcgatc gtcacatcgatc ggcgcacccgg gacgtatcg
tcgcccacgtt ggcgcacacg ctgcgcgcgc gccgcacaggaa gttcccgatc cgacgcgc
tcgeetgcac cagccgcacc gatgcgtatcgatc cctgtgtcgatc cgcgcggac gccaaacgt
tcgcccacgg cgacggccaa tccggcgcgc tgccgcgc gtcgtatcgatc tccggcgccg
gcaacgcgtatc cgcggatcgatc ggcggggatg tgctcgatc tccggcgatc ttccggcgatc
aaacccggatcgatc tccgtgtcgatc tcccgatcgatc tcccgatcgatc tcccgatcgatc
cggttccatcgatc ggacgtcgatc cggcgccgcgatcgatc tcccgatcgatc tcccgatcgatc
ccgcgtctt tgctcgatc tccgtgtcgatc cccgtatcgatc gatgggtcgatc ggtgtcgatc
ccgacgggtt catcgatcgatc agcatggggc agtacacggc cgcgtatcgatc gccggcgatc
tgctcgatc gacacgcgtatc ggcgtatcgatc tcccgatcgatc tcccgatcgatc tcccgatcgatc
ccggccggggat gatgtcgatc tcccgatcgatc cccgtatcgatc gatgggtcgatc ggtgtcgatc
atggactcgatc gatcgatcgatc gtcgtatcgatc cccgtatcgatc gatgggtcgatc ggtgtcgatc
acgcgtatcgatc ggccttgcg cgcgtatcgatc tcccgatcgatc tcccgatcgatc tcccgatcgatc
acatcgatcgatc cgcggcgatc tcccgatcgatc tcccgatcgatc tcccgatcgatc tcccgatcgatc
tcccgatcgatc tcccgatcgatc tcccgatcgatc tcccgatcgatc tcccgatcgatc tcccgatcgatc
cggttccatcgatc gacacgcgtatc gacacgcgtatc tcccgatcgatc tcccgatcgatc tcccgatcgatc
ccgttccatcgatc tcccgatcgatc tcccgatcgatc tcccgatcgatc tcccgatcgatc tcccgatcgatc
aagtccggcc cggacacggatcgatc tcccgatcgatc tcccgatcgatc tcccgatcgatc tcccgatcgatc
ccgcacgtatc tcaagcgatc tcccgatcgatc tcccgatcgatc tcccgatcgatc tcccgatcgatc
tgcacgcgc ggtggcgatcgatc tcccgatcgatc tcccgatcgatc tcccgatcgatc tcccgatcgatc
acagaggcgatc gggcaggatcgatc tcccgatcgatc tcccgatcgatc tcccgatcgatc tcccgatcgatc
acctggatcgatc atcgatcgatc tcccgatcgatc tcccgatcgatc tcccgatcgatc tcccgatcgatc
ccgatgtatcgatc tcccgatcgatc tcccgatcgatc tcccgatcgatc tcccgatcgatc tcccgatcgatc
atcgatcgatc tcccgatcgatc tcccgatcgatc tcccgatcgatc tcccgatcgatc tcccgatcgatc
agatcgatcgatc tcccgatcgatc tcccgatcgatc tcccgatcgatc tcccgatcgatc tcccgatcgatc
cgagtctcgatc gttcaaggatcgatc tcccgatcgatc tcccgatcgatc tcccgatcgatc tcccgatcgatc
aggcaccgcgatc tcccgatcgatc tcccgatcgatc tcccgatcgatc tcccgatcgatc tcccgatcgatc
ttccctcgatc ggcggaaatcgatc tcccgatcgatc tcccgatcgatc tcccgatcgatc tcccgatcgatc
gcccctcgatc ggcctggatcgatc tcccgatcgatc tcccgatcgatc tcccgatcgatc tcccgatcgatc
gcccctcgatc ggcctggatcgatc tcccgatcgatc tcccgatcgatc tcccgatcgatc tcccgatcgatc
ttccaaacagatc attcgatcgatc tcccgatcgatc tcccgatcgatc tcccgatcgatc tcccgatcgatc

tgttgtccccgc	gatgccggccg	ccagcgtcg	ccgtcgccgc	ggccgacgcg	ggactgaagt	29760	19/35
cccgcgctac	gacggcgccc	gtctcgaa	cgcattggc	ttcaaggccg	atcgacggca	29820	
aagccagtgc	ggagtcgtcg	cccggcaga	tggcccgct	cggcccttgg	atcgatcgat	29880	
acacacggcg	cacggccgga	tcaaggcgc	gcctggccga	gtaccggccg	gtgcttgcgg	29940	
atccacgctc	ggtcggccgga	ttcaaccggc	tcttggaa	gatgatctac	ccgatcatca	30000	
ccgaccggtc	ggaaggatcg	aaggcttggg	acgtcgacgg	caacgat	gtggatctcg	30060	
tgggggggtt	cggtcgctt	ctgttcggac	accgttcgc	gatcggtggc	gacggcggtgg	30120	
aggaacaact	ccaccgggggg	ttcgagctcg	gaccgttgc	gcccgtcgcc	ggcgaggatcg	30180	
cggcgctcg	caaggaattc	acggggatgg	agecgcttgg	attctccaac	accgatcg	30240	
aagccgtgct	ggccggccacg	cgaaaa	gacgggtcac	cggggcgcc	aaatcgccg	30300	
tcttcgaagg	gtcgatcac	gttctccctg	acaatgtct	cagggggccg	ctcggtgtga	30360	
acggcgaatt	gcggteggcc	cccggggcgc	ccggcatcgc	cggcagcgc	gtgagcgaag	30420	
tcatccgtct	cgagta	aaccggaa	cgctcgagg	catccgcgc	cacgatcg	30480	
agatccgcgc	gatccctcg	gagccgtgc	agagccggcg	gctcgacct	cagccgaagg	30540	
agttccgtca	cgagggttcgc	cgcttgggg	acgagatcg	cgccgcgtc	atcttcgac	30600	
aagtcatcac	cggttccgc	gtcata	gcggcggccca	ggcgacact	ggcgcccg	30660	
cggtatctgc	cgctac	aaggctcg	gcggcggcat	gcccattcg	atcgatcg	30720	
gcaatcgaa	gttcatggac	gcgttcgac	gcccggact	ggagtagcgc	gacgggtcg	30780	
tcccggaggt	gggggtc	tactttgc	gcacgttcat	gcccgcatt	ctcgatcg	30840	
ccgctgc	ggccgtgtcg	acgtacat	agtcgcagg	tcccggct	cagacgcggc	30900	
tggcagatcg	gatcgagcg	ctcgac	acgatcg	gatcggtggc	aggacacggcg	30960	
cgccgtatca	gattac	ttcagctcg	tgatgagct	gaacttcc	cacgaccaga	31020	
agacggcggt	gtgtgtt	ttcctgat	gcaatcg	gattcatatc	tggagggcc	31080	
gtcccttctt	cttcac	gtcac	aggccacta	cgccgcatt	ctccgcgc	31140	
tcgacgatcg	gctggccg	atcg	ctggcttctt	cgccgcgccc	gcccact	31200	
cattgcagg	gaccgggac	gcccggagg	acgttctgc	gttca	ggacagcagg	31260	
agatctcggt	cgccggcc	atggggc	cgatcac	cgccataca	gaatcg	31320	
cgctcgat	gcccggcc	ctgaat	cgatcg	gcccgcgt	gacgagg	31380	
tcgatcgat	cgatcg	cgat	tgccagcg	accgttgc	accgtc	31440	
ttccgggac	gtcgg	gttgaat	aggacgt	gggatt	gaggacgt	31500	
agcacgaggc	cgatcg	gtgctcg	gcccggac	gtggcg	gacgtggac	31560	
gcggggccgtt	gtcg	accaccat	cgatc	cgatcat	gttctcg	31620	
tggcggccca	ccatctcg	tgtgac	gttgc	gtcattct	cgcatct	31680	
cgacgttcta	ttcg	gttca	ggcgt	cctcg	ccatgc	31740	
tgagccgtt	cgcg	gatcg	gatcg	gttca	ggcagagacc	31800	
aggcgttctg	atccgg	ttcg	accac	gtcg	tttacgac	31860	
ggccgccc	gcca	tcc	cgac	gttgc	tttacgac	31920	
cgatcgatcg	cgcc	gatcg	accac	gttgc	tttacgac	31980	
tgctccgg	cttc	cgatcg	actcg	gttgc	tttacgac	32040	
gcatccgg	cgcc	cctcg	cgatcg	gttgc	tttacgac	32100	
tcgtccgg	tcgc	gtgt	atcgat	gttgc	tttacgac	32160	
tgaggacgca	gtca	ctcaat	gatcg	gttgc	tttacgac	32220	
ccaagctcg	gttgc	gttca	cgatcg	gttgc	tttacgac	32280	
tcgagccgc	catcg	ccccgg	cgatcc	gttgc	tttacgac	32340	
cgcgctg	atcg	atcg	cgatcg	gttgc	tttacgac	32400	
ttgtcgaggc	cgat	accat	cgatcg	gttgc	tttacgac	32460	
atagctatcg	gatct	gatcg	accat	gttgc	tttacgac	32520	
tgccgggtat	ctcg	gatcg	accat	gttgc	tttacgac	32580	
ccgactaccc	cgcc	gtc	atcgat	gttgc	tttacgac	32640	
cgaggggact	cgcg	tgt	cgatcg	gttgc	tttacgac	32700	
gcccggaa	ttgg	gtc	atcgat	gttgc	tttacgac	32760	
cgatcgatcg	gttgc	gtc	atcgat	gttgc	tttacgac	32820	
gatcgatcg	gttgc	gtc	atcgat	gttgc	tttacgac	32880	
ccagggaccc	gttgc	gtc	atcgat	gttgc	tttacgac	32940	
atgatctgaa	gttgc	gtc	atcgat	gttgc	tttacgac	33000	
cgacgggaca	accg	gtc	atcgat	gttgc	tttacgac	33060	
cggtatcg	acgg	gtc	atcgat	gttgc	tttacgac	33120	
cgcccgat	cggt	gtc	atcgat	gttgc	tttacgac	33180	
tcacccatcg	gttc	gtc	atcgat	gttgc	tttacgac	33240	
tcgacttct	gttgc	gtc	atcgat	gttgc	tttacgac	33300	
cctcgctgt	gccc	gtc	atcgat	gttgc	tttacgac	33360	
tggaaagact	cgac	gtc	atcgat	gttgc	tttacgac	33420	
agaccacgg	cggt	gtc	atcgat	gttgc	tttacgac	33480	
gctgtccggc	cgcc	gtc	atcgat	gttgc	tttacgac	33540	
cgcccgcc	gatt	gtc	atcgat	gttgc	tttacgac	33600	
ggtatctcaa	tcg	tc	atcgat	gttgc	tttacgac	33660	
agccgggagc	gccc	tc	atcgat	gttgc	tttacgac	33720	
tcgagttct	cggt	tc	atcgat	gttgc	tttacgac	33780	
	cggt	tc	atcgat	gttgc	tttacgac	33840	

gcggaaatcga	agccggcgctt	cgcgctcatg	aggcgctcg	gcaggccgt	gtcgtgctgc	33900	20/35
aaaaatcga	gccggcgac	gagcgctga	tcggctatgt	cacgaccggc	ggcgccgcct	33960	
cgatctcgat	ggcgagctt	cgcacgtcg	tcaagcagag	gttgceggac	tacatggtgc	34020	
cgccgggtat	cgtcgcgctc	gacaagctgc	cggtcacgtc	gattggcaag	atcgatcgac	34080	
gcccctcc	ggccccggcc	gagcgcgcca	gttgcagtc	ggagttcgtc	acgcccggca	34140	
ccgagacgga	gcccggctt	gcccggatct	ggagcgcggt	gtcgggtctc	gagaagattg	34200	
gcccgcacga	caacttcttc	gacctggcg	ggcactcgct	gtcggggcc	cgcaccggca	34260	
tgaaggctcg	cgacacgttt	cgcgtggacg	tgtcgctcat	cacgttcttc	gaaatgccc	34320	
ccatcgccat	gatggcgccg	gcaatcgacg	ccgttaccgc	ggccctgtcc	tccggacca	34380	
tcacgcccag	ggcccgctc	agggcccgca	gggatgagct	tttgcggcca	tgaacccccc	34440	
ccaccccgag	actgagtccc	ccgacctcg	gcccgcgg	atcgccggc	aaggccggca	34500	
cgtgtttctg	gcccggcggt	cgtacggcca	ggggcgctt	ttgttcttgg	cgcagctcg	34560	
gcctgaaagc	tccctctatc	acatccgat	cgcgtcaga	ctgaagggcc	gcctcgaccc	34620	
cgccgcgtc	gaacgcgcca	tcaacaccat	cggttccagct	cacgagacgc	tgccgacgac	34680	
gttctcgctg	caggcgatg	agctcaagca	ggtcatactct	ccgacgctgc	gtattccgat	34740	
ctccctcg	gacatcgatc	tgctcagtc	tgagaaaacgc	gaggccggat	ggccgcgggt	34800	
tgcggccgac	gacggcgccg	gctgttctcg	tctggcttag	ggcccgctgc	tcaaggctcg	34860	
actcgtcatc	ctcgagccct	ccgatttctg	ccttctgtatc	accgtgcata	acatcgcttt	34920	
cgacgggtgg	tccggcgaaa	tcttcatccg	cgagctggcc	gaggcgta	cggtcgagct	34980	
cgacggggcgc	gaccgcgaa	tgccggagct	ggcgatccag	tacgcccact	ttgcggcggt	35040	
ccagaaggaa	cagaccgaaag	ggcccggtct	ggccgtgtcg	tttgcggact	ggacgcagga	35100	
gctggcgaaa	gctgtcgatga	cgctcgacct	gegacgtcg	cgatctggc	cgccctgtca	35160	
gacctcccg	ggccgcctcg	cctcgttc	gtctcagcg	gcatctggc	cgccgtcg	35220	
cgcattcage	cggggcgaga	acgcgacggt	gttcatgacg	cttctcttgg	ccttcggcg	35280	
cgtgtcgat	cgtgggtcc	gacaggaaaga	cattctatc	ggcagcccg	cgccggccgg	35340	
cgatcgccg	gagcttgagg	gcctgtatcg	gttcttcatc	aacacgttc	tcctcggtct	35400	
caaccgctcg	ggagatccca	cgtttcgca	gttcgtcggt	cgggcgcc	gcgcgtgcgg	35460	
tgcggcttac	gcccgtcagg	acctggcg	cgatcggtc	gtcgagacgc	tcaacccgga	35520	
gcccggccgg	gatccggatc	cgatctatca	agtcatgttc	gcccggcg	cgccggcg	35580	
gaagatgacc	ttccggcgaaa	tcgagctcg	gcccgttgc	gtccacaacc	gcagcaccaa	35640	
gtgcgacctc	acgctcgagg	ttccggaaaga	cgccgacggc	gtcaccctgt	acgtcgagta	35700	
cagcgctc	ctgttcgatg	cggcgacggt	ggcgtcgctg	ctgcageaat	atcgccgtgt	35760	
gtttcagacc	gcatcgccgg	acccggacca	gcccgtctcc	gagctcgctc	tgcttggcga	35820	
cgacgacgca	atcgacgtca	tcgcacttgg	gaccgggtcc	gcccggctct	acccggcg	35880	
cgtcgatctg	gccaacggct	tcgaagacg	ggtccggcg	actccggcg	cagtggccgc	35940	
gacgctggag	ggacacgacc	tcacgtacg	ggatctcaac	cgccaggcg	acccgttgc	36000	
ccggcggtt	aaggcgatcg	gtgtggccc	gaacgtcg	gtcggcg	cgctcgaccg	36060	
gtcgttcgat	ctgttgggtt	cgcttctcg	cggttgcgaa	gcccggcg	cgtatctccc	36120	
gctcgatttc	gctcaccggc	aggaacggct	tgcttcatg	ctcgccgaca	cgccggcg	36180	
catcggtc	acgctcgccc	gatttccgg	ggcgtcgag	tcctttcccc	tcttctgcet	36240	
ggacgacgag	ctggggggca	ccggcgagga	gcccggacaa	gatctcg	cgccgtcgac	36300	
cgagagaccac	ctggccatcg	tgatgtac	ctcggtcg	actggtc	ccaaaggccgt	36360	
cgcgatcatc	cattcgccgg	tggttctgt	cgatcggcgc	acagactacg	tccgtggaa	36420	
cgacgtcg	gcatcgatcc	acatcgcc	gacgtcg	gacgcgttga	cgttcgaggt	36480	
gttggcgccg	tttgcgtacg	gcccggct	ggtgcgtgtt	cccgagccgc	tcgtcagcct	36540	
ggagacgctc	gagcgacgc	tccggtcg	gcaggtggac	tgcctgtgc	tcacgaccgc	36600	
gcttcaac	gcccgtatcg	acgcaagcc	cgccattctc	gtcggcgta	agcaactct	36660	
gatcgccgc	gaagcgctgt	cagtgcgca	cgtgcggcgc	gcccggcg	acctgcccga	36720	
caccggtc	gtgaacgcgt	acggtccgac	cgagtgcacg	accatcgct	gcccgtacca	36780	
gattccgc	acgctcgatc	cgcaggcg	gtcatttc	atcgccgc	cgattgc	36840	
tafcaggcg	ctcattctcg	atcgacacgg	cgaccctcg	cggttgcgca	tcggggcg	36900	
gctctgcctc	ggcgagacg	ggctggcg	ggaataccgt	aatcagccgg	cgctcactgc	36960	
ggaacggtt	gtgccaacg	cgttcagcc	cgagccggcc	gccccggct	acccgtacgg	37020	
agacctcg	cggtacggcc	gcaacgcg	catcgatgc	tttgcggg	tttgcgtacca	37080	
gatcaaaatc	cgccggatcc	ggatcgac	gggcgagatc	gagcatggcc	ttcggtcccc	37140	
gggagacgtt	ctggacgcgg	tggttggat	ccgggtcgat	gcccacatgg	cccacatctgt	37200	
cgcctacgtc	gtgcgcggcc	atgggagtca	actaccgg	accgacgtgc	gcgaacgtct	37260	
gaaggcgcc	ctgccccgaa	actgtgtcg	cgcgaagatc	gtgttcttgg	cccgaggccc	37320	
gcaacggcg	gcccggcaagg	tgcattcg	ggcgtcgct	gcatcgccg	tcgaggcgcc	37380	
cgcggccgag	ccgcgcgatc	cgtcgttct	gaacgagg	gaagcgccg	tcgcgcgaaat	37440	
ctggagcgaa	gtgttccggc	tgccctccgt	gtctcgac	gacaattttt	tcgatcg	37500	

Figure 5b

```

CDS complement(3..914)
/blast_score=2e-66
/blastp_match="AE004644.PA2177 Pseudomonas aeruginosa"
/gene="regulator hybrid"
/note="probable similarity to prokaryote sensory
transduction proteins"
/product="sensor/response regulator hybrid"

CDS 924..2168
/note="none"
/gene="ligase"
/blastp_match="AP003013.MLR8297 Mesorhizobium loti"
/blast_score=e-152
/product="2-amino-3-ketobutyrate CoA ligase"

CDS 2207..3190
/blast_score=e-151
/blastp_match="AE008872.TDH Salmonella typhimurium"
/note="none"
/gene="dehydrogenase"
/product="threonine 3-dehydrogenase"
/pfam_match="PF00107; adh_zinc; 1"

CDS 3373..4455
/note="putative"
/gene="methyltransferase"
/blastp_match="AE001866.DR0026 Deinococcus radiodurans"
/blast_score=1e-08

CDS 4546..4959
/blast_score=2e-11
/blastp_match="AP002997.MLL1617 Mesorhizobium loti"
/gene="unknown"
/note="pfam00263, GSPII_III, Bacterial type II and III
secretion system protein, Expect = 7.8"

CDS 5176..6192
/blast_score=5e-98
/blastp_match="AF064070.PE20 Burkholderia pseudomallei"
/gene="glucose epimerase"
/note="putative"
/product="UDP-glucose 4-epimerase"
/pfam_match="PF01370; Epimerase; 1"

CDS 6331..14043
/note="substrat AT, malonyl ; zinc depend dehydrogenase ;
zinc dependent adenosine deaminase putative"
/gene="PKS I"
/blastp_match="AF285636.WCBR 2547 Burkholderia mallei"
/blast_score=0.0
/pfam_match="ketoadyl-synt"
/pfam_match="ketoadyl-synt_C"
/pfam_match="Acyl_transf"
/pfam_match="SAM binding"
/pfam_match="adh_zinc"
/pfam_match="pp-binding"

CDS 14275..15408
/blast_score=e-104
/blastp_match="AF285636.WCBT Burkholderia mallei"
/gene="acyl-CoA transferase WcbT"

```

```

/note="putative"
/pfam_match="PF00155; aminotran_1_2; 1"
CDS
15436..16245
/blast_score=5e-11
/blastp_match="TTDEFFMT.FMT T.thermophilus"
/gene="formyltransferase"
/note="evidence experimental"
/product="methionyl-tRNA formyltransferase"
/pfam_match="PF00551; formyl_transf; 1"
/pfam_match="PF02911; formyl_transf_C; 1"
CDS
16287..17384
/note="putative"
/gene="glycotransferase"
/blastp_match="AF285636.WCBD Burkholderia mallei"
/blast_score=e-99
CDS
17427..18158
/blast_score=2e-82
/blastp_match="AF285636.WZM Burkholderia mallei"
/gene="ABC-2 transporter Wzm"
/note="putative"
/pfam_match="PF01061; ABC2_membrane; 1"
CDS
18248..18847
/blast_score=7e-61
/blastp_match="AF285636.WZT Burkholderia mallei"
/gene="ABC-2 transporter Wzt"
/note="putative "
/pfam_match="PF00005; ABC_tran; 1"
CDS
18952..20346
/note="putative"
/gene="glycosyltransferase"
/blast_score=e-101
/blastp_match="AF285636.WCBE Burkholderia mallei"
/pfam_match="Glycos_transf_1"
CDS
20442..21167
/note="putative"
/gene="unknow"
/blast_score=1e-26
/blastp_match="AE009248.ATU3189 Agrobacterium tumefaciens"
CDS
complement(21164..24301)
/note="tranporter domain"
/gene="unknow"
/blast_score=5e-29
/blastp_match="AE009122.ATU1658 Agrobacterium tumefaciens"
/prosite_match="PS00402; BPD_TRANSP_INN_MEMBR; UNKNOWN_1"
complement(24351..27023)
/note="none"
/gene="unknow"
/blast_score=5e-29
/blastp_match="AP003581.ALR0267 Nostoc sp"
complement(27806..29686)
/note="none"
/gene="cell surfarce protein"
/blast_score=2e-34
/blastp_match="AE010748.MA0851. 2567 Methanosaarcina
acetivorans"
/product="cell surface protein"
complement(29535..30872)
CDS

```

```

/note="none"
/gene="cell surface protein"
/blast_score=2e-36
/blastp_match="AE010748.MA0851 Methanosarcina acetivorans"
complement(30848..32647)
CDS
/note="fragment"
/gene="O-antigen"
/blast_score=1e-72
/blastp_match="AF105060.RFBC Riftia pachyptila"
/product="O-antigen biosynthesis protein"
/pfam_match="PF00535; Glycos_transf_2"
complement(32574..35555)
CDS
/note="putative"
/gene="glycosyltransferase"
/blast_score=7e-46
/blastp_match="AE013462.MM2213 Methanosarcina mazei "
complement(35533..36598)
CDS
/note="putative"
/gene="glycosyltransferase"
/blast_score=7e-37
/blastp_match="AE013462.MM2213 Methanosarcina mazei "
complement(36516..37400)
CDS
/note="putative ATP/GTP-binding protein, esterase fush
9e-14"
/blast_score=8e-22
/blastp_match="AP003581.ALR0267 Nostoc sp"
/gene="ALR0267"
Sequence 37507 BP; 5531 A; 13507 C; 13011 G; 5458 T; 0 other;

```

gatcccacga	tccggcgccag	cggattgcgc	agctcgtcg	cgagcatcgc	gaggaactcg	60
gtcacgtggc	ggccctcgcc	ctccagctcc	tccatgcgt	tgcgtgcga	gagggtcg	120
gtcaccttgg	cgaagcccat	gtgccttcca	ctcgctcg	aaagcgccgt	cacgacgacg	180
tccggccaga	agcgctgtcc	gtccttgcgc	acgcgcacgc	ccgtgtcctc	ggcgccggcg	240
tgcagcagcg	catcgtgcag	ctcgccctcc	gggcggccgg	ccgcggcata	ctcgccggga	300
tagaagagcg	aaaagtgcgg	tccgatgatc	tcttcggcgc	tgtagccgg	gatgcgtcg	360
gcgccttgt	tccagctgg	cacgtggccc	gggggttcca	gcatgtat	cgcgtatcc	420
tgcacgcct	cgatcaggag	gcgcgagcgc	tccctcgctt	ggccgactg	ctctctgtgc	480
tgcggcgct	cggtcagggtc	gcgcgtgacc	ttgccgaagc	cgaccagcg	gccgtctcg	540
tgcgcagcg	ccgtgatgat	gacgttgccc	cagaagcgcc	ggccgtcctt	gcgcacgcgc	600
cageccctgt	cctcgaagcg	gccgtcctcg	cgtgcgcga	gcagtcctt	ggcggggccag	660
tgcgcgcga	tgtcttcctc	ggtgtagaac	gcccggat	gatctcttcg	720	
gcfgatagc	ccttcaactg	tccggccccc	cgccggat	gatcttcacg	780	
tcgagcacga	agatcgctg	gtccgtgacg	gaccggacca	gatggcgtcc	840	
tcgcccggca	ggccgctcg	ggcgccgaag	gcggaggtgc	gcacccggaa	900	
gatgggcttg	gcatggctat	taggtacaa	cggtgcacgc	aagttgcaat	960	
gagataattg	tttgcatttt	atgcatttt	tggaaatttt	acgcaggag	1020	
tcgaggccca	gggcctgtac	aatccgagc	cggtgctcg	ctctccgcag	1080	
tgcagggtgc	ccccacgac	gtgcctcaact	tctgcgcctt	ggcgccacgc	1140	
accacccggc	cctggctcg	ggcgcaagg	agacgctcg	ggcgacggc	1200	
cgtcggttcg	tttcatctgc	gggacgcacg	cggtgcacgc	tacggcatgg	1260	
cgcgttccct	ccgcacccgag	gacgcgtatcc	cgagctggag	acgcgcctcg	1320	
ggctttcga	gacgctgtcg	ggcgaggagg	tgtaacggctc	ctgcttcgac	1380	
cctcgatcat	cgacggcggt	ggccctctcg	acgcgggtat	gccaaacggcg	1440	
acatggcgag	cctcgaggcc	cagctgaagg	aggcgaagcg	cttccgctac	1500	
tgatcgccac	cgacggcggt	ttctcgatgg	cgccgatcgt	gacaacaacg	1560	
gcgagctggc	cgggcgctac	cacgcgtatgg	tgatggtgg	cgactcgcac	1620	

tcatgggcgc	gcacgggcgg	ggaacgcccgg	agcattgcgg	tgtcgaaagg	aagggtggaca	1680
tcctcacccgg	cacgtcgcc	aaggccctgg	gccccgccttc	cgccggctac	acggcggggca	1740
agcgcgaggt	ggtgggcctgg	cttcgcaacc	gctcgccccc	ctatctttc	tccaaacacgc	1800
tgtatgcccg	catcgccggc	gctcgctca	aggtgctcga	tctccctcgag	ggccggccgg	1860
agctcgccgc	gaagctcgcc	cgcaacgccc	gccacttccg	cgccgagatg	acgcccctcg	1920
gtttcacgt	ggccggccgg	gaccatccga	tcatcccggt	gatgtgggc	gaggccccgc	1980
tccgcaagga	gatggcgac	cggctgtga	aggaaggcat	ctacgtggtc	ggcttctcg	2040
ttcccgttgt	gccccgaggg	caggcgcgc	tccgcacgc	gatgtcgcc	gccccatgaac	2100
cgaagcacgt	cgatcgcc	atcgccct	ttgccaagg	cgccgcgcgt	ttgggagtca	2160
ttgcttgaga	accctttcga	agacgaagcg	cgagccggc	atctgatgg	tcgagtcgccc	2220
caagccgtg	gtcgccaca	acgacgtgt	gatccgcgt	aagaagaccg	ccatctgcgg	2280
caccgacatg	cacatcttca	actggacga	ctggcgcag	aagaccatcc	cggtgcggat	2340
gacggtcggc	cacgagta	taggcattgt	ggaggccatg	ggccaggagg	tgccggcct	2400
gcagggtcg	cagcgcgtc	ccggcgagg	ccacatcg	tgccgcatt	gcccgaactg	2460
ccgcgcggg	cgccgcacc	tgtgcgc	cacgcagg	gtgggcgt	accgcgggg	2520
cgccgtcg	gactacctgg	tgattccgc	ggagaacgc	ttcccatttc	ccgacgacat	2580
ccccgacag	atcgctcga	tcctcgatcc	gttcggcaac	gcccgcaca	ccgcgccttc	2640
ttcgcacgt	gtgggcgagg	acgtgtat	caccggcgc	ggccgcata	gatcatggc	2700
cgccgcattc	gcgcgcac	tcggcgc	ccacgtcg	atcacggaca	tgaacgacta	2760
ccgcgtggc	ctgcgcac	agatgggtc	cagccgcgc	gtgaacgtat	cgaaaggaaaa	2820
cctgaaggac	gtgatgcgc	agctcgcat	gttcgagg	ttcgcgtcg	gcatggagat	2880
gtcgggcgtg	ccctcgcc	tccgcagat	gctcgacacc	atgaaccacg	gcccgaagat	2940
cgccatgctc	ggcatccgc	ccagcgagg	cgcgatcg	tggacccagg	tgtatctcaa	3000
gggcctgtc	atcaaggcg	tctacgg	cgagatgtt	gagacctgg	acaagatgtat	3060
cgccatgctg	cagagtggc	tcgac	geccatgg	acgcaccg	tcgacgtgc	3120
cgagta	ttcgac	agacgatgg	ctccgg	tccggaaag	tcgtcctgtc	3180
gtgggattag	cgagggccac	ggatcg	ccgg	cgccggcgt	cgtgcaggcg	3240
ggttggcagc	acgccccgg	atttcgc	tggtc	cgacttcgc	cgtcacgc	3300
ccgatcgca	tcgacgtc	cgtgac	cg	atgcccagg	cacctcgctg	3360
tggacgcgt	cgatggcc	cagcgg	gac	ggctacc	cgccacgtat	3420
cgcccttcc	tgcccttgg	gccgttcc	ttcc	cgccgcctc	gttcgagttc	3480
gctctcg	tgccgggtc	agg	cg	ggcgttgg	tgatcttgc	3540
cccgccgg	cacccac	ctggc	cc	ccaccgc	ggcgttgg	3600
acgcggc	tcgagac	ggccttgg	gat	ac	cttggatggc	3660
ttcgaccatg	cgtcgc	gtcg	tac	ccggatgg	cttggatggc	3720
gggcgtgc	tggatgc	atgcgg	ggc	cgctcg	cgccgtgc	3780
tatgc	cccg	agg	cc	g	gttcccac	3840
atcctcgagc	gccacgc	cg	gg	taac	gttcccac	3900
gcccgtcca	acgcgtccc	gttgc	gg	cc	ctcggttcc	3960
gcccgtgc	atttcgtgg	cggt	cc	cg	ctcggttcc	4020
aagccggcg	gcctgttgt	catccat	cc	cc	ccacggccat	4080
cacccggcg	agtacagcc	cgac	tt	cc	ccacggccat	4140
cgcgacatcg	tgttgcgc	ggac	cc	cc	cgacgagtg	4200
cgccgcgg	actggcgt	gtaca	cc	cc	cgccgcac	4260
gacgagttc	gcgcgtc	cttc	gg	cc	cgagccgt	4320
gtggagtaca	agccggag	gtcg	cc	cc	ctccgagctc	4380
ta	gtatcaac	caag	cc	cc	agctggcgt	4440
gcccattcc	gttaagg	ggc	cc	cc	cttacaggac	4500
cgtcactcc	ggactg	gat	cc	cc	gcatggcag	4560
caggccgtc	tcg	c	cc	cc	ttgttatccc	4620
gcccgtc	tgac	gag	cc	cc	gacccagacc	4680
ggcacgttgg	tcgt	gagg	cc	cc	ccgcgcggaa	4740
accgc	ccac	gg	cc	cc	ccggccatg	4800
ccgttgcgt	agacc	ccac	gg	cc	cgtggcgtc	4860
gcgcagtgc	ccgc	gg	cc	cc	gatgaccgag	4920
aagggccggcg	agg	ccac	gg	cc	cgaggccac	4980
aagccgcgc	aagg	ccac	gg	cc	gattcgaacc	5040

ttcgacccccc	tgcaccccat	gcaggtgcgc	taccaggctg	cgctacgccc	cgaccgaact	5100
tgaaattata	cctgcategg	ggtcgcagcc	gccatgaatc	gaaacgcccc	ggggttccggg	5160
tagcatccgc	ccatggcgac	gaccctcg	acgggtgg	caggctacat	cgccacccac	5220
atcctctcg	cactggccca	ggccgggcga	cgcagcatct	gcatcgacaa	ctactccaac	5280
agctcgccgc	gttccatcga	acgcgtctg	cagatcgac	ccggttgcgt	cgaggccotc	5340
gacgtggaca	tccgcgacgc	cgacggcata	cgcaaggta	tgcagggcg	cgacgtggac	5400
agcgtgatcc	acctcgccgg	actgaaggcg	gtcggcata	cggtcgagca	gcccggagcgc	5460
taccacgaca	acaacgtccg	cggcacccgag	agcctgtctg	cggcgtcgc	cgactcgccg	5520
gtgcgcaagt	tcgtttcag	ctcctcgcc	accgtgtacg	gcctcgccga	gaagatgccc	5580
atcgacgagg	acgcggccac	ctgcggcgag	agcccgtacg	gccagaacaa	gctcgacatc	5640
gagcacatgc	tggtcgcct	cgccaagcgc	gaccctct	ggcgcgtagt	gaacctgcgc	5700
tacttcaacc	cggtcgccgc	acacgagagc	gcgcgtatcg	gcgaggaccc	ggccgggggtt	5760
cccaacaacc	tcatggcta	cgtgtccag	gtcgcctcg	ggcgaactgaa	ggagctgagc	5820
gtctacggca	gcgactaccc	cacccggac	ggcaccggca	tgcgcgattt	catccacgtc	5880
tgcgacccgg	ccgaaggcca	cgtggccg	ctggaaagc	tgcagcgcgc	ggccccccggg	5940
accgttcata	cgtgtaatct	cgggacgggg	ctcggctaca	ccgtgtcgca	attgtcgaa	6000
accttegagc	gcgtgaacaa	actaagcatt	gcccggcgca	tgcgtggggcg	gcggccgggc	6060
gacgtcgca	tctgctacgc	gaacgcgggc	cgcgcaagg	ccgtgtcg	ttggacccgcg	6120
aagcgggaa	tcgaggagat	gtgcccgt	gcctggcg	ggcaggaaag	aatgcacaaag	6180
ggctacgggt	gacggattgc	aaatcatggc	cttggcgacc	tttgctgca	tgcggcatga	6240
agggggcggt	tcttgttaha	tccaaagg	gggatgcgtc	caggacgtca	aaaacagcac	6300
ccggcggttgg	aaggaagaag	gcaaccctag	tgttgtcatg	gctccagacg	ggattgcgcg	6360
ccgagtaaga	atcgctagtc	gagaccaact	cgccctccggg	tgcaacgaat	cgtgagccga	6420
agcaaacacct	cgcacgcgt	cgccgtcg	ggcattgcgt	tccgccttc	cgcgctacg	6480
cgcgaatccc	tctggccctc	gtcctcg	gggcgcac	tgcgtaccac	cgtcgatccc	6540
gcgcgttgg	cgaaggagcc	gtactaccac	ccgcgcac	ccgagccgg	caccagctac	6600
tgcgtcccg	ccggcaccat	cgcgacatc	agcggctcg	acgcgtcg	cttcggcata	6660
tcccccgcg	aggccgcgca	gatggatccg	cagcagcgc	tgctgtcg	gctctctgg	6720
gaagccctcg	aggccgcggc	ggtgaagccc	tcgaccctcg	gggcgcgca	ttgcggcg	6780
ttcgtggcc	tgtccacccgt	cgactacagc	tacggcctcg	cctacgac	ggcgtccatg	6840
gacgcgtgt	cggccaccgg	caacaccgcg	agcatcg	ccaaaccgc	ctcgatattc	6900
ta	cccgactcc	acggccccc	catggcgt	gatacggcg	gctcgctc	6960
ttccaccagg	cctgcccgt	gatcgccatc	ggcgagacgc	gccaggcg	ggtggccgc	7020
atcagcctgc	acctgcatec	gtacgggttc	atcgcttct	ccaaggcg	gatgctctcg	7080
aagcggg	cctgcccgt	gttcgacg	agcggcgat	gta	ctcgaaaggc	7140
gcggg	ttgtgtcg	ggaccc	ccgcggcg	ccgacggcaa	cccgatctc	7200
gcgtcg	ccgcgacgg	cgtaa	ggggccgt	cctccagc	caccgtgccc	7260
agcgtcgat	cccgacgtcg	attgtcg	gatgtgtac	agcgcgcgg	catcgacccg	7320
tccgagatcg	actacctcg	ggccac	accggcacc	tggggcg	tccgatcg	7380
acccgcgc	tcggcatcg	gtcg	ccgcgc	ccaccc	gtcgccatc	7440
ggttcgg	agagcaacgt	cgccac	gaggcg	ccggcatgg	gggcctgg	7500
aaggcgctgc	attgcctcaa	gcac	gtgc	ccatccac	cgatcgccc	7560
aaccccaaca	tccacttcg	cgagt	ccgg	ctga	gtcgac	7620
catccgtaca	agaagctcg	gtcg	ccgg	ccgc	gtcgac	7680
cacgtatcc	tcgagagc	gccc	ccgg	ccaa	gtcgac	7740
gtggcgctca	tgctgac	ccac	ccgg	aggcg	ccatccac	7800
gcgaacttcc	tgcaggcg	cgag	ccgg	ccgc	ccatccac	7860
ca	actggcac	gtc	ccgg	ccgc	ccatccac	7920
gcgg	tcg	ccgg	ccgg	ccgc	ccatccac	7980
gcgtcg	cccg	ccgg	ccgg	ccgc	ccatccac	8040
gcaatgggc	gcgg	ccgg	ccgg	ccgc	ccatccac	8100
gacgcgtgt	tccggccgt	cg	ccgg	ccgc	ccatccac	8160
cccgccgacc	agctggac	cac	ccgg	ccgc	ccatccac	8220
ggcttacccg	agctgtcg	ccact	ccgg	ccgc	ccatccac	8280
gtggcgagg	tcg	ccgg	ccgg	ccgc	ccatccac	8340
gtgatctacc	accgcagcc	ctacc	ccgg	ccgc	ccatccac	8400
gtcgccctcg	ccgaggac	ccgc	ccgg	ccgc	ccatccac	8460

gtgaacgtcg	cgtgcataa	cagcacgcgc	aacgtgacgc	tgcgggcac	gwgccggc	8520
atcgaggcgc	tcgaggcga	gtcaacgcag	cgcaagggtgt	tccacgcgc	cctcgaccc	8580
gactacgcgt	tccacagccc	ggcgatggac	ccgggtgcgc	acgggctcg	gagcgcgctg	8640
cgccggccta	cgccgcgc	cacccgggtc	gcgttccatt	cgccggtcac	cggtcgccc	8700
gccggggca	accagctga	cgccaccc	tgttgagca	acatccgaga	gccgggtgcgc	8760
ttccaggccg	cgatccgcgg	catcgccag	tccggcgtca	atgtttcat	cgagatcg	8820
ccgcacccga	tcctcaagaa	ctacatcaac	gacggcctgc	gcccgcctc	gatcgaggc	8880
cgcgcgctgg	tcacgctga	gctcgccgc	agcgatcg	cgccgatcc	cgccgcgc	8940
caggaagtcg	tgatcacgg	ctgccccgt	gagaccgcga	agatgtccc	cgcgaggc	9000
cacttcgtcg	agtcgcgc	ctaccgtgg	cagcgcgagc	gccactggcg	cgcccccacc	9060
tcgcaggcct	acgacccat	ccagcacggc	aagcagcacc	cgctgctggg	ctatcgctg	9120
cacgagaacg	acttcagtg	ggaaaaccac	atcgacaccc	cgcttatcc	ggcgta	9180
gaccatgtcg	tcggcgccgc	cggtgttgc	cccgccgcgc	gcttgcgtca	gatggcgctc	9240
gcccgcctcg	cgatcg	gggcccgcac	gogcacgaga	tgcagacgt	cgagatccgc	9300
agcgcgctgc	tgctcgagga	ctccacgtcg	aagaccgtgc	gcttcgcgt	gaaaccgc	9360
ggccgcctca	caatccgcag	ccgcgc	ctgagcgagg	atccgtggca	actccacgt	9420
gtcggcaagc	ttgtcgccac	gcccacccg	cttccgcgc	cgccgcgc	gtcgccgc	9480
tcgcgc	ccgatgtgat	cgccgc	caactacgaa	gcccgcgc	ggccgcgc	9540
gcctacgc	cgcccttcca	gtcggttcc	aagggtgtggc	tgcacgc	ggaccggg	9600
agcgcc	cgcccttgc	cctcccaag	ccgatcg	gctcgatggc	cgatgcac	9660
ctgcaccccg	cgtcgttgc	cggtgttcc	cagctgtgg	tgcacccgt	cgccgcgc	9720
gcctcccg	acgcgcagat	cgccctcg	ccgatctgg	tgcggcgac	gcccctctac	9780
ggccctgc	ccctcg	cgccctgc	gtccgcctga	ccgcgc	cccgcgtcg	9840
ctggtcgc	acttcgagct	ctacggcccc	aacggc	tggtgcgc	gtcaccggc	9900
gtcgcttc	gcggcgtca	gctgaagc	ccgcagg	ccgcgc	ctacctgtcg	9960
taccgc	ccccgcct	gaatggc	gacacgc	gcccgc	gctgcgc	10020
gcgttgc	ccgcgcgt	cgccgc	tggatcc	tgcacgc	ggccgc	10080
cgctactacg	atgaggtcg	gctgcgc	tacccg	aggccgc	gctgatgc	10140
gcgtcg	agoggatcg	cgaggcccc	ctatcg	ccgcgc	cgccgcgtcg	10200
acgcggcc	agcattggcc	gtcctcaag	cgcgc	gatgtcg	cgaggacc	10260
ttgtcg	ccgccc	cggtggc	tggagcc	tcaccg	gcccgc	10320
caggagag	ggatcac	gctgcgc	tacccg	aggccgc	gctgatgc	10380
ctcgccgc	ccggcgtca	cctcg	gtgtgc	gcaagg	gtccttgc	10440
agcctgttgc	cgccacgc	ggcga	tgcgc	tgtgtcg	cgccgc	10500
ttcgcca	ccgggcgc	catcg	acgctgg	tgcgc	gcccgc	10560
gcccacccgc	cgctcg	gctcg	acgcgc	gtccgc	cacgcgc	10620
gtcctcg	ccgtggac	cgacc	gagtacc	tgcgc	cacggac	10680
gcactggcg	agtacga	gtcctcg	gccattcc	acgtcg	ttcg	10740
gacctgaa	tccgggcat	ggcc	gcccga	aggccgtt	cgatgtcg	10800
atcgttcc	atggcctg	cggt	gatccg	ccgcgc	gcaccc	10860
ggcg	cggaagacgg	gctgcgt	atggtgc	agcatcc	gctgg	10920
gacctct	tccggcgg	ccggcgt	tggac	ggcc	ctcg	10980
tcgcgc	gttcgcgg	cgagt	gtcg	ccacctegg	ttcg	11040
accgtcg	tggccg	gcccgc	atcg	ccgcgc	catggc	11100
cgcgacggcc	cggtgcgc	ggct	gagcgc	cgatcg	gggtcg	11160
ttcg	aaagcg	ttatcg	gctgtcg	cctgcgt	ggcggat	11220
cgcgacgc	gcccgcgt	gttc	gagat	ccgtcg	ggccgagat	11280
gcccacgc	ccgcgc	ccag	gcgc	ccgcgc	ttccctgg	11340
ggccctcc	acgacgc	ccgc	gacat	ccgcgc	cttcctgg	11400
gcccgc	ccgcgc	gccc	gcgc	ccgcgc	ccgcgc	11460
gcccgc	ccgcgc	gccc	gatgc	ccgcgc	catgac	11520
gcccgt	ccgcgc	gccc	gatgc	ccgcgc	ccccgcct	11580
gcccgt	ccgcgc	gccc	gatgc	ccgcgc	ccgcgc	11640
gcgttgc	ccgcgc	gccc	gatgc	ccgcgc	ccgcgc	11700
gcgttgc	ccgcgc	gccc	gatgc	ccgcgc	ccgcgc	11760
gtggcctca	ccggcgcc	ccgc	gatgc	ccgcgc	ccgcgc	11820
tccgtcg	ccgc	ccgc	gatgc	ccgcgc	ccgcgc	11880
aagaaccc	cgatggcg	ccgc	gatgc	ccgcgc	ccgcgc	

gaggtgcgcg	ccggggcct	caactccgc	gacgtgatgt	acgcgatggg	cctgctctcc	11940
gacgaggcg	tcgaggcgg	ttcgccggc	gcctcgctcg	gcatggagct	ctcgccgcgc	12000
gtcaccccg	tcggcaagga	cgtgacctcg	gtcgccgcgc	gacgaggt	gctcgcttc	12060
gcgcctcg	cgttgcac	gcacgtggc	accaccgcgc	attcggtcgc	gaagaagccc	12120
gcccgttgc	cgttgcgtc	cgccgcacc	atccccacc	cgttcttcac	cggtactac	12180
tcgttgaag	acctggccca	gctgcccgg	ggcgagcgcgc	tccgtatcca	ccgcgcgcgc	12240
ggcggcggt	gcacgcgc	gatccaggc	gccaagtggc	tggcgccgga	gatcttcggc	12300
acccggcgt	ccgcagagaa	gacgcacttc	gtgcgcctgc	tggcgccgga	ccacgtgc	12360
gactcgcc	cgctcacgtt	ccgcgcac	gtgctgcgc	tcaccaacgg	cgaggcg	12420
gacgtgggc	tgaactcgct	ctccggcgg	gccattgcgc	gcaacctgcg	cgcgctgcgc	12480
ccgttccggc	gcttcatcg	gctcggaag	cgcgactact	acgagaacac	gcacatggc	12540
ctgccccgt	tccgcaacaa	catctcgat	ttcggcgtcg	atgcccacca	gctcatgaag	12600
gagcgcccc	acctcgccgc	ccagcttcc	aacgagctga	tgcagctt	cgagcagg	12660
gtgtctcg	cgctgcctca	ccgcgccttc	gccgcccaccc	aggcggtgga	agccttccgc	12720
tacatgcac	actcgccca	gatcggaag	gtgggtctgt	cgttcgccga	cgcggtgaag	12780
gcgcagccg	cccgccccgc	cggtccgc	gagctggcgc	tgtcgccgaa	cccccacgtac	12840
ctcgtaacc	gcccgttgc	gggcttcggc	ctgcgcaccc	cgcaatggct	gttgggacaag	12900
ggtcccccc	acctgggtgt	ggtctccaag	ageggcgcc	agtcgctcg	gaccaaggcc	12960
gccccggcg	acctcgaggc	ccaggccgc	acgggtateg	cgaggcg	cgacatcacc	13020
gaccgcgcct	ccgtgcaacg	gctgctcgcc	gaggctcg	ccgcgttgc	gccgtgcgc	13080
ggcgtatcc	acgcggcg	cgtgatccag	gacggcttca	tcaccaacat	gacggccgc	13140
cagatcccg	acgtgctggc	gcccaaggc	cteggcgc	gccaccttgg	cgagctcacc	13200
cgcgccgca	agctcgattt	cttcgtctg	tactcgctg	cgacgacgt	tttcggcaac	13260
ccggcccg	ccaactacat	cgccgcaac	tgcttcctcg	aggcactcg	caaggcg	13320
cgcgctcg	cccgccggc	gctgtcg	ggctggggcg	cgatcgggg	cgtgggtac	13380
ctcgccgc	acgagaaggt	gaaggaagcc	ctgcagacgc	acatgggg	caccgcac	13440
gagtccgg	ccgcgtcgc	ggtgctggag	cagctctgc	tcgcgcac	ctcgccctc	13500
ggcgtctcg	acttcagctg	gwgcaac	cgccgttcc	tcccacccgc	cgccctcg	13560
cgctcccg	agctggcg	gwgccgcac	gacgcgaagg	acgacgac	ccggcg	13620
gagctcgcc	gcctcgccaa	cgagctggac	gcctccg	tctcggcgat	tttcacccgac	13680
ctgctcgcc	gwgagggtgg	cgagatcctg	cgcatccgc	cgaggccct	cgatacg	13740
cagcccttgc	aggaaatggg	catggattcg	ctgatgggg	tggagctgt	gaccgcgg	13800
gaggcgcc	tcggcgtgaa	cctgcgggt	atgtcgct	ccgagcagcc	gtcgatcg	13860
aagctggc	accgcac	gwgccgt	aaggacccg	acgcgg	cgaggccgaa	13920
tcgcccac	accagatcg	gwggtcg	gwgccgt	cgcccg	cgactcg	13980
caggctcg	agctcac	ggcgtcg	gacgcga	ccacgc	gggacgtcc	14040
tgaaggcg	caaccccg	ggcctctcg	aggcgtgaa	ggagcag	atccagcg	14100
tgctcgac	gwggtcg	cggtcg	aggacaagg	ctcgccg	gagtccg	14160
tcaacgcctt	ccggccgac	cccgatgccc	ttggccacat	ctcgagcg	tactcg	14220
tcgacatcg	ccggctctac	cagcagatcg	agctggtgg	ggaaagg	gcaagctgg	14280
gcattcccg	tccgtacttc	cgcgccac	acggcg	cgccgat	acgcgcatt	14340
ccggccgca	gtacatcaac	ttctccag	acaactac	cggttctcg	ggcgacccgg	14400
ccgtggatgc	cgccgccaag	caggccatcg	acgaaatacg	gaccctcg	ctcgcc	14460
aggctcg	ccggcg	gwgccgt	cgccg	cgccg	cgccgagg	14520
tatggcgtc	acgcgc	ggccttcatc	agcggcc	cgaccaac	ctccacatc	14580
ggccacctgt	tcggccg	cgacccatc	gtgcac	cggtcg	caacacatc	14640
ctgatggc	tcggcgt	gggccc	cgcatgg	tccgcacaa	cgactgg	14700
gccctggac	agctcg	cgccact	cgccact	cgccact	gatcgatc	14760
gagggcg	acagcatgg	ccgggactac	cccgg	ccgaaattctc	gwgccctcg	14820
cgccgc	gwggttcc	gatggtgg	gwgcc	ccttcgg	gatggcc	14880
cgccgc	gwggttcc	gatggtgg	gwgcc	gatgg	catctggat	14940
ggcgc	gwggttcc	gatggtgg	gwgcc	atggccg	cgaggccg	15000
ggcgc	gwggttcc	gatggtgg	gwgcc	ggccgttca	tcggccgg	15060
gtcgagc	tgaagttcg	cgccccc	ttctctaca	gctcg	ggccccccc	15120
gtggccg	ccggcg	cgccgt	cgcatcg	cgccgt	ccgcgtgg	15180
gccctgc	ccggcg	cttcttc	gaggccg	gwgccg	catcgac	15240
ggccttc	ccggcg	cggtgt	gwgcc	gwgccg	ccgcgcgg	15300
cgccctgg	ccggcg	ccagcg	atcaacgtcc	agccgat	ttacccgg	

gtgcaggaga	actcgccg	cctgcgttc	tgcgtgagcg	ccacgcacac	cgaggagcag	15360
ctgcgttca	cgtccgcga	gctcgccgac	gcctggcga	agctctgagt	ggcgcccc	15420
cggctgcgcg	tgcgtctg	cacccacggc	gccctgaacg	gcfgcgttgt	gtctcgccg	15480
ctgctggccg	cccccaegct	cgaagtctcc	gcfgctggta	tctcagccg	cgcgccggt	15540
gccccatgagt	cgtatggccg	tgccgcgctc	gatacgtgc	gcfgagcgg	cgtcgcctat	15600
gcfgctgtacc	tctggtgcgc	cacggcgctc	gccgacactc	tgcgtgcegg	cacgtccgag	15660
ggccccggc	cgcgcatcg	cctgcgcgc	gcatccgc	tcctcgccac	gcccggcggt	15720
aacgatgcca	ccgcgcgcgc	tttcatcg	ggggcgccgc	cgacactgtat	cgatcccg	15780
tttttcaacc	agacatcg	cgccgacgt	gctgcgctcg	cgcgctggc	cgcggtaaac	15840
atccacccgt	cggcgctgc	gcaactccgc	gcfgtgatc	cggtgtccct	cgcgccctg	15900
cgcgggtccg	agcgccacgg	cgtgagcgt	catcgcatcg	aaccggctt	cgataccggc	15960
gcfgctgtcg	cccaaggaaac	cgacgtcg	gccccggca	gcfgcttgc	cgccacccgc	16020
gcfgctatg	accgcggggc	ggcgctc	gcccggccgt	cgccggcgct	ggccggccac	16080
cccgccggaa	ccccccagcc	ggggggggc	tcctacgact	cctggccac	cccgccccag	16140
gtgcggcct	tccggcgcgc	cggggggc	ctgcgtccgc	cccgcgac	gtggcgctc	16200
gcgcgcggg	ggccggccgc	tttctaata	gaatcagcgc	gtagcgccg	actcacgc	16260
cgctttccc	cgacccccc	aggccatga	aacactggct	gaagcaacac	gcaatctcg	16320
cgctcaccgt	ctgtcgccg	acgtggccg	cgatcctca	tttgcgcctg	atgcgcctcg	16380
acgtctacat	ctccgaatca	cgcttcgt	tgaggagccc	ccagcgccag	gtgcagaccg	16440
gcctgggg	cgcgcgtc	tcgggcac	gttctcg	ctcccgaggac	gacacctact	16500
cggcgcacga	cttcatcacc	tcgcgcac	cgctggcga	gctggacaag	aagctcgcc	16560
tccgcaagct	ctacacggcc	gccaacatcg	acttcatcaa	ccgcttccc	gggctcgact	16620
gggacgacag	cttcgaggcg	ttccacccgt	actaccagaa	gaaggtcag	atcgacttcg	16680
acacccgtc	ctcgatc	gtgctcg	tgcgcc	cgagaaggcc	gactcgccg	16740
gcatcaacga	cctgctgt	cagatggcg	agcgcctgt	gaacgagctg	aacgagcg	16800
gccgcacca	cctgatccgc	ttcgcc	ccgaggtc	gctgcgc	gacaagggt	16860
aggatggcgc	gttggcgtc	tccgc	gagcaacca	gtcggtt	gagccgacc	16920
gccaggcctc	gatccagct	caggcg	ccaagctca	ggaggagct	atgcgcacc	16980
aggggcagct	ggcacagct	cgcaagct	cgccgcacaa	ccgcagatc	ggcgccgtcg	17040
agaacaagtc	ggcggcgct	cgctggc	ttggcgcgoga	gtcgccaag	gtgaccggcg	17100
gcagcggtc	gttcagcg	cgcgc	cggtc	actaccgt	gagaaggct	17160
tcgcccac	ccagctggc	tttgc	ccgcgt	cgacgc	agcgaggc	17220
agaggaagca	gtctac	gagcgc	tgca	cctgc	aatcggtcg	17280
agccgcgc	cattcg	atcttac	ttgt	gttgcgt	gggcttgg	17340
tggtggact	gttggggc	agcgt	gac	ctaggcc	gcctgggg	17400
cttcccttc	gcgcgc	gggttgc	ggcgcgt	ctacgc	ctgatcg	17460
aggtcatcac	gcgtt	cgcac	tcggcgt	gtgg	gtcgaggcc	17520
tgtatctcac	gtcg	acggc	tttgc	ggatcg	cacggctcg	17580
cgctcgat	cg	gggtt	cacc	gatact	tggcgaact	17640
gcgcctcg	cg	ccatcg	gg	ccgtcttc	caccgcac	17700
tgcgcgtat	cg	acgtgtt	gt	ctac	ccacggcg	17760
cgttcaccgt	gt	tcgttgc	gtt	atcg	ccgatcgaca	17820
tgctcaaggt	cg	ttcg	gttgcgt	gggcgc	tggcgtcg	17880
ccatcg	cg	cgatcg	cc	ccgcgt	ccgacggc	17940
acctgcgtt	cccg	tcgt	cc	atcg	gagaaattcc	18000
agcagtctgt	gt	tcgttgc	cc	gtcg	gacggctact	18060
tcggcaacgc	gg	tcgt	cc	gtcg	tcgtggcc	18120
tgcgcgtct	cg	tcgt	cc	ccgcgt	ccgcgt	18180
ccgggtcg	cg	atcg	cc	atcg	ccgcgt	18240
cagttcg	ctcg	tcgt	cc	gtcg	ccgcggcaa	18300
gtccacgct	atcg	tcag	cc	ccgcgt	ccgcgt	18360
cgagatgagc	gtgt	ccgt	cc	atcg	ccgcgt	18420
gctggacaac	ctga	atcg	cc	gtcg	ccgcgt	18480
cttcgtcg	atcg	tcgt	cc	ccgcgt	ccgcgt	18540
gcacggcat	at	tcgt	cc	atcg	ccgcgt	18600
cctgatcgac	gagg	ccgt	cc	gtcg	ccgcgt	18660
gctttccac	aagg	ccgt	cc	ccgcgt	ccgcgt	18720

caagcttat	tgcgagagcg	cctgcgtgct	gcacgagggg	cggctgctgc	cttccccac	18780
cgtggacgcc	gcctacgagt	tctacatgaa	cgaggtcatg	caggacctcg	cccccggaggt	18840
tgcctgaacg	cccccgcgcg	gcccgtgcgc	gtgcgtctcg	agctgcccgc	ggcggttcgac	18900
ggccacgcgc	gcatccccca	ggaggcacgc	ctgcgttcc	gcccgcgtcg	catgatcgag	18960
ggcatcgacg	tggagggcct	gctcagcac	agcggccacg	tgctcgc当地	gggcctgccc	19020
ccgcgcggcg	gcccggacct	cgcacccggac	cgccagctga	accgcctgtc	gcccgtcg	19080
gtctccacca	agcaggagct	caccaacgcg	cacgtggcg	cggcgctgat	ggccattcgc	19140
cagctgctgg	gcccgcgcga	gacccttac	cgcttcgacg	cggcgactt	ccgcgacttc	19200
gtctggcagg	cgcttccgc	gcccacgtg	cacgcccggg	acttcgactc	ggtcaccggc	19260
gcccggcttc	gcccggcg	cggtccgtgg	acgggcatgc	accgcgtgc	gctggtcacg	19320
cgcaagctcg	gctacacgcg	gcatccgcgc	atcgacactt	cgcatttcga	cgtgcgtgatc	19380
gcccggacgc	cctacccggc	gcccgtcg	gacggcacgc	gcctcgtggt	gcccgtaccac	19440
gacgcgatcc	cgctcctcat	gcccacacc	atctccgaca	tgctctacca	ccaggccctcg	19500
cactacccgg	cgctgcgcgc	caacgtggcc	tcgggcgcgc	acttcgctg	cgtctccgaa	19560
gccacgcgca	aggacctgt	ctcggtctt	cccgagggtcg	aggcgcgctc	cagcaccatc	19620
cacaacatgg	tctcgaccca	ctacttccgc	gagaccgcgg	acgcggggcg	catcgaggag	19680
atccctgcgaa	cgccgcgcgag	cgagcgcata	aagggcgcgc	ccaacggagc	cgccaccgccc	19740
gtcgcacgat	cgcccttcctt	cctcgcgcgc	gccgcgcacac	ccagcgaccc	cggctacactg	19800
ctcgcgtgg	cgacgatcga	gcccgcgaag	aaccacgcgg	gactgatcgc	cgcgtggag	19860
cacetcgcga	cctcgcgttt	cccgggcctg	cgccctgtgg	tggtcggcat	gcccggctgg	19920
ggetacgagg	cgctggtagc	caagttcaag	ccgtggctcg	cgccgcggca	gctcttcgtg	19980
etcgaggagc	tgcccggcccc	cgagctgcgc	ctgctctact	cgcacgcgcg	tgccacgggtg	20040
tgcccggagct	tcggcgaggg	cttcgattt	tccggcgtgg	aggcgatcg	ctgcggcage	20100
ccgggtatcg	cctcggagat	cgccgcgcac	cgcgagggtgt	accgcgcacgc	cgccgagatc	20160
tgcaagccct	attccgtggc	cgacctcgcc	gaggcgatcg	ggcgcgtcat	cgacccggcg	20220
gcaacggggcc	tgccgcaggc	gctggtcacg	cgccgcacccg	aggctcgcg	gcccgtacaq	20280
cccgaaggcca	tcctgcgcga	gtggcgcgag	tacctgctcg	gcacgggtcg	cgcggaggcc	20340
ccgtgagcga	tccgaaggcc	accgaccagg	cgtgggagga	gtggggccag	cgcgaccct	20400
acttcggcgt	gatcacaac	ccgcgttcc	gccgcggggca	gatggacgag	gaagcgaagc	20460
gcaaattect	cgcgccggc	cgcgccatg	ccgactacgt	gatgcgcatg	gtgcacgcgc	20520
acatcgcgccc	cgacttcgt	ccgcgcacca	tcctcgattt	eggctgcggc	gtggggcc	20580
tcgtcatccc	cttcgcgcgc	caggccgagc	aggtgaccgg	cgccgcacgc	tcccccctcg	20640
tgctggccga	ggctgcgcgc	aactgcgcgc	agcaggggcgt	ggcaaatcg	cgcctcgtgg	20700
tctcgacga	ctcgctcacc	gggctgcgcgg	gccccttcga	cctcgac	tgttccatcg	20760
tgttccagca	catcgatccg	gcccgcgggc	gcccgcgtt	cgccgcgcct	ctggggcacc	20820
tggcgagcgg	cggcggtggc	cggttgcact	tcgtctatgc	gaagcgtata	tacgcggcga	20880
cgtacggcgt	gggcgcaccc	cccgcgcgc	cgccgcgcgc	gcccgcgcgc	cagcgttccac	20940
ccagccgggc	ggagatcaag	gcccgcgcgc	aggcgccggc	cgccgcgcgc	acgcaggcgc	21000
gggcgcggcga	accggcacc	aatccgcacc	ccgacatcg	gatgaattcc	tacccggcgg	21060
cggagatgt	gttcctcggt	caggaagccg	cggtcacgcg	cttcacgc	gagttcacc	21120
accacggcgg	cgagctgggt	ctttccctct	tcttcgc	gcccattggc	tccttcgagg	21180
cgaaggagag	gcccagcgcg	gtgatecg	cgccgtgtcg	cgccgcgcgc	gcccgcgcgg	21240
ccgcgcacttc	ggcagccac	cgctcgagg	cgccgcgcgc	cccgta	cgacgcgcgc	21300
agttagtgcac	ccagtaggag	tccggacagg	ccgacgcgc	tcccg	cgctcgac	21360
cgggcccgac	gcccgcctgg	tgcacttgg	cgccgcgcgc	cgccgcgcgc	acggagggt	21420
cgaggggcag	gacggcgac	agcgcctcg	tcacccgc	cgccgcgcgc	tcatcgccgt	21480
cgtcggcgag	gcccgcgc	gtgggcagg	actccggcgc	aaggcggtgc	gcccgcgcga	21540
gcaaatcgcg	cagcatcgca	gcccgcgc	cgccgcgcgc	caggcgccgc	cggtggtagt	21600
gggtccagaa	cgaaaggcgcc	gcccgcac	cgccgcgc	gacccgcgc	cttgcgtcg	21660
ccccggagag	ccgggcacgc	gcccacaaca	ggttcgc	cgccgcgcgc	gcccgcacgc	21720
tccgcaacgg	cgccgcaccc	gcccgcgc	cctcg	accctcg	gacgagaagg	21780
ccgagcgccgc	gaactcgatc	agcgattcg	cgagcgac	atgttcggc	cagagccgc	21840
cgggatccac	cgccgcgc	gcccgcgc	cgccgcgc	cgctcctcg	tcaccgcgaa	21900
gcccgtgcac	ccgcaggcg	tggaaacgc	cccactcg	cgacatcg	gcccgcgcgc	21960
gctccagcgc	accctcgac	gcccacatcg	tgcgc	cgccatgtgc	aggatggccc	22020
aggtcgcctg	cgccgcgggg	aggttcg	cgccgcgc	gagcgggtcg	aggaccgcga	22080
gcccgcggc	ggcgctcg	gcccactg	atcgcttggc	caggccgatc	agcgcgcgc	22140

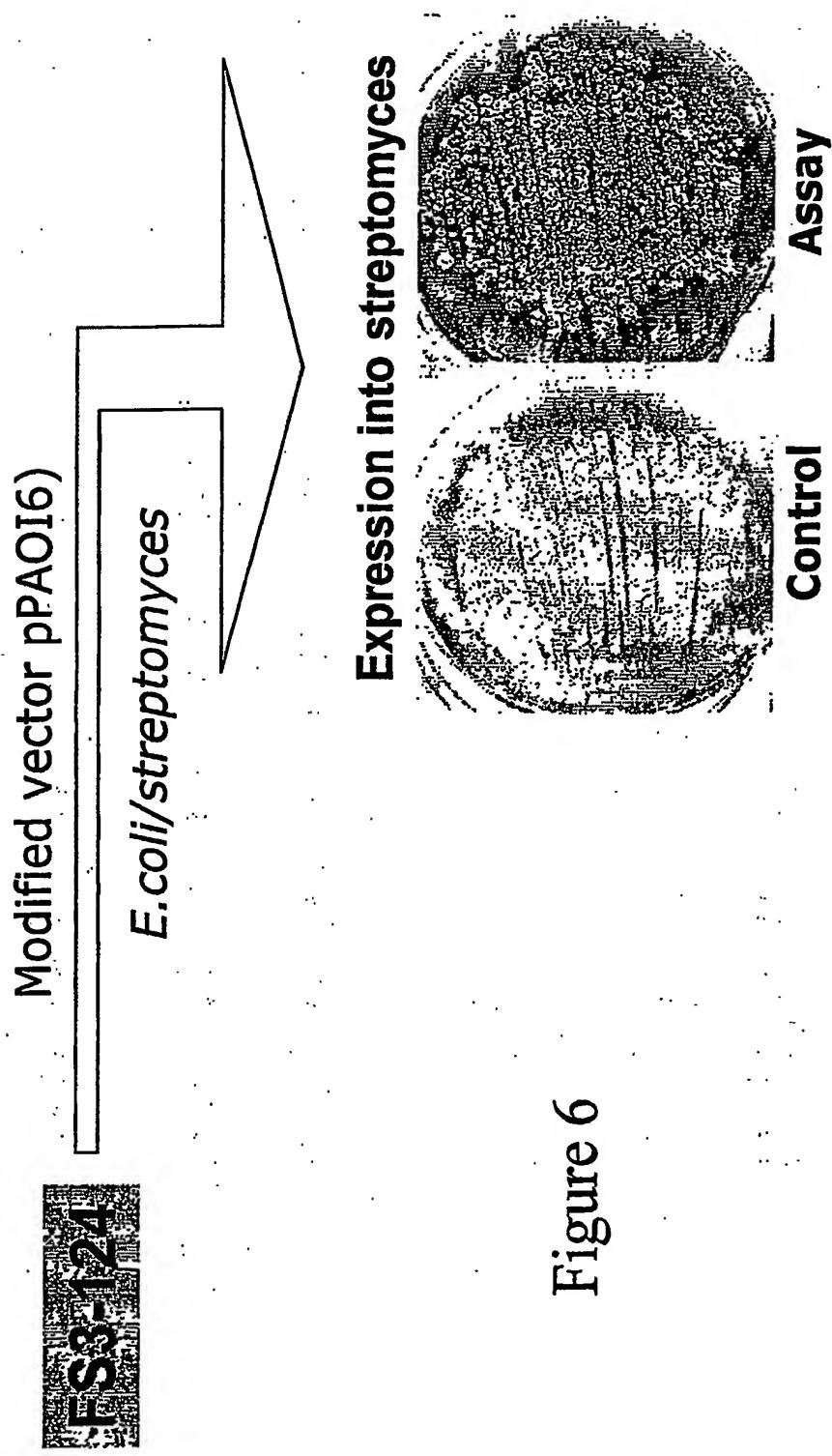
cgagctcgaa	ggccgcgccc	aaggcggccc	ccgcgtcg	ccccaccgcc	gcgagggtcg	22200
cgcgcagccg	gtagaggcgc	gcccgggtgg	aatagcccca	gaacgggtcg	gggctcgca	22260
tcaagtgaatc	gaacgagcgc	tcgacactcg	cgcacccacgtt	gcccacgtt	gcgagggtcc	22320
cgcggccacag	ggtcgcccga	tgcgcggcca	tcggctcg	ggggacgaag	gcgcgggtgga	22380
tcaagctcgac	gtcgccgaa	cgcgcgtcg	gcaacgcgag	ctccaggcg	ttcagccgct	22440
cgaggctcg	cttgtaactcg	cgataggcat	cgccggtcg	gtcgatttcc	tgcgcggcac	22500
tgtcgccat	cagcgaggcc	tgagccgt	gttgcggcg	aacgggtcg	tatgcgttagg	22560
ccgcgcgatc	gtcgtagggc	agcgtggccg	cgccgcgaga	cccgaacgcg	gcgcggtcg	22620
ggaatccag	cttgcggcag	acgcgggggt	ccaggtcg	ggcttccatc	tgcggccggc	22680
tccagtcg	cagcgccg	acgttcg	agagccacaa	cgccgtgcgg	tagagctcg	22740
ccaggtccgc	gtccggggcg	ccctgcgcct	cgaagatcg	cttccactcc	gcgagtcct	22800
cgtggccg	ggagaagatc	ggcacgtgg	cgccgaagat	cgccagcacg	tagggccg	22860
cgcgcacactc	gagacgtcg	accagcgac	gcagcgcgc	gtcggttcc	gccatcaccc	22920
cgcagttactt	gtcgatcg	ccctggcg	cgtgcccgg	gaggaagtgc	ggcgcacagt	22980
agcgcggc	gggcgcctt	tcgcccagg	agggtccgt	gcccatac	gagacgcgca	23040
acacgaagcg	cggccgcgaa	tgcgtcg	ccgcgtatcac	ggcatccatc	acggccgaaat	23100
ccgcgtatcg	ctcgaaactgg	cgcggcgat	ggcgaacgc	ctcgatcg	tggAACGCCG	23160
cgaagccgag	cttttgtac	acgcggtcg	gttgcgtgg	ccagcgggtgg	aacggatgg	23220
tggccacgct	gcggtagccg	tggccgc	gcatgcgcg	caccgcatac	acccggtcg	23280
gcaactggcc	catgtgggc	agcgcgcgg	gccttcctc	gccaatgcg	atgcgggtga	23340
ggatctcgaa	ctcggtcg	cagggttgc	cgccgatcac	ttcacgcgc	aggcggccgg	23400
tggtgcggcc	cggccgcgtc	gcccggagga	acggcagcgg	gtcgatcg	agcggcaca	23460
cgaacggccg	caggcggcg	acgcactca	tcatcaggaa	gatgtatcg	ggcgctccg	23520
cgggtggccgg	gcccgtcg	gagcgtcc	ggcgctcc	cagccgc	acgcctcg	23580
gcagccggcc	cggcgcgacc	gggcgggcg	tgctccggc	gaccacgc	gcgtcagga	23640
cgaagcccg	gcggcgacg	ttcacgtt	gtcgggtgg	cgaatacgtc	atcgatcg	23700
actccacccg	ccagcggcgc	aagcgtccg	cgacgaaggc	gacgcgc	gcggcgagca	23760
cgagcagcca	ctgcccggc	tcgtcg	cgatcg	gtcgatggc	tcgggtatct	23820
cgaacgcggc	ccacgccc	gccgcgagca	ctgcggc	ggcaacgcg	cgcaccgca	23880
cgcggggcc	atgcaggta	gccggcg	acaccacgc	ctcgatcg	tcaggaaat	23940
cgcggggcc	gaccacgcg	ccgaagaacg	cgagcaggcc	tgcgatcc	gcgttagagcg	24000
cgcggcc	cgcacgg	cccacccgt	cgagcgc	gaccgcgac	aagacgcg	24060
catggaaat	ggcgagcagg	ccggcg	aggccagg	ccgcgcgc	gccacgggtt	24120
ccacgtata	gcccaccacg	aggcggcg	cgagggcg	ctgcagcacc	gcgttaggcg	24180
gcacgcagac	cagcg	ccagcagca	cgccgaggaa	acccgc	ccagccgtgt	24240
cgtcgaggat	catgggacca	ccagggacgt	ttccggca	gtctataacg	ccccggccca	24300
tgaaaaacgg	cgaggccgg	gggcctcg	gtggagggaa	ccgcgc	tcagggttgc	24360
gggggtgcg	ccgaccagcc	gcccggcc	atcagggtt	cgctcg	ggtggcgagc	24420
ccgttcatca	ggcgcaccgt	gacgctcg	tcgggttggc	ggaagaccag	gtccttgcgg	24480
ctgtcgccgt	tcagggtcc	cagctcg	acgttccaa	ccgtgc	cggcagcgc	24540
tcgcccgcgg	tgtatgtt	aacaccatcc	atcaggcg	tgtgcacgc	gcgtcggt	24600
tggcgaaaga	ccaggtcg	ttcgcgt	cggtcg	cgccaccag	ttcaccgac	24660
cacccgtcg	cggccggcag	gagctggcg	ctggcgcc	aggtgttgc	gttcatcagg	24720
aacagggtcg	cgcggccgtc	ggtgtggcg	aagatcatgt	ccgcgcgtcc	gtcgcgcgtt	24780
aggtcgccca	ggtggcttac	cgtccagcc	ctggcgcc	agagaaagcc	cgctccggcc	24840
gtcaccgtgg	tgcgtccat	gatgtatgt	tagccgc	cgtcgtgt	catgaagacc	24900
aggtccgcct	tgcgtcg	ttcatgtcg	ccgccttgg	tgagcgttca	gcgcgtcccc	24960
gccccgaaga	gctggcg	gcccgtat	gtcgatcc	ccatcaacca	gatgtgcgc	25020
cgtccgtcg	tgtggcg	gatgtatcc	gccttgc	ccgcgttcc	gtccgcgtt	25080
tggctcacga	tccagccg	cccgccggc	agcagtc	tgcggccgt	caccgtatc	25140
ccgttcatga	tgtacgcgt	ggcgcggcc	tcgggttggc	ggaacaggat	gtcgatcg	25200
aggtcgccat	ttaggtcg	cgtctgg	accgtcc	ccgcgcggc	gcgcgtatcagg	25260
ttggccgagc	cgtgtatggc	ggtgcgtt	atcgatcc	ccgcgtatgc	ccgcgtcg	25320
ttctgcagca	cgcggcg	cctgcgt	ccgcgtatc	ccgagacgc	ccgcgttgg	25380
ggcggcg	cgcacccg	aaccggc	gtccagcc	cgatcc	gcccgtgg	25440
cagtggcg	catccggggc	cggcggtt	caggcgg	agttcaagg	cgtggagagg	25500
aagttcgaga	aacgcccac	cagcgtgt	gcgcggc	tcaccgg	cgacacc	25560

ttccagcccc aggccgtggt gccccgcgtcg ctgaacacca ccgtgccgac gatcaggaac	25620
tgcgacttgc cggcggtgg gaagctcgac acgcgatagg cgaacgcgac gttcccgcg	25680
agcagecggtgc cgcgaacga gaggtcgag ttacccgtcg tcgagttggt cgggtcgctc	25740
gtcaccacct gcccgaacg aaggctggtc gcgcctcga aggctggtc ggagccgacc	25800
gtccatgcgt tggcccgcc cgaggtcggt aagccggcag gcagctgccc acccggtggc	25860
cacggcgaga ggtatgtgcc ggtcgccgtc gccggggccgc cgatgaagac acccggtggaa	25920
cccgtcgcat tggagagcgc gacgtgaac gtctcggtgc cttcacccgt accgtcgctc	25980
gccaccggga ccgtgtggt cttggcggtg gtctcgctg tggcccgact gagcgagccc	26040
gaggtcgccg ttagtcggc gcccgaagt ggggtgcgt tggaggtggc gtagctgacc	26100
gagatggcaca cgcgcgagcc gcegtgcgg ctcacggta gcttcacgtt gccggcggtt	26160
tcccgcccg cgaagggtggt gccgggtgaac tgcaccgaac cgggaacggc gggcggttg	26220
accgtggact tgagcgccga gagtggcg cggttgggt tcagcgagag cgcgtcggtt	26280
gcctgcgcgg tggcgacgg ttgcggcga ccgacaccccg tcgtgcacga gaggcccg	26340
ttggagaagc ggtagacctt ggtcgccgt gtagtcgag tcatgcagca ggcacatgat ggtgtgtaa	26400
tccggccgt ctttcgtcg gcactcggtt cgctggctca cgcgcgaggc gccgcgcgg	26460
gagccgttgg gcaggggccga gttcagggtg aggttgcgc tcgcgcataa gaagtggccg	26520
tacgatgtacg cgaacgcacc ttgcggcggtt ctcgcgggtc cgcggcttc ccaggcctgc	26580
gagggccggt cgtggcggtt gcccattggca tggcccgact cgtggatgaa gacccactcg	26640
caacccaaca ggcagccgga agcaacggcg taggcaaagt tgggtccgg cgtgtcggt	26700
ccgatccagg ctgcgcgtt gcccggaaag tccggccgt tgcgcataa ggcacccatg	26760
tccgeccgtt attgcgtcg gatggctcg atgttgcgcg acgtcgccgc gtcgaagctc	26820
gcgtgcggcgg ggggtatggc atgcatacgcc gtgggtcggt cgtggcggtc gtcgtagttc	26880
acctgcgtcg cattcaccag ggcgcgcgtg atggcgaccc cgctgtcggt atacgcactg	26940
ttggcgcccg tgacgaggaa gttcaggcgc gtcatcaggc tgcgcgcgg ggcgtggcc	27000
aagccgcgcg agtacacgat catcaggcg atgacgttct gggcggtgg cgtcacctgg	27060
gcgagccgac ggtatttcacg ccggggatgg cgatcagggt ctggggctg cggcgcgcga	27120
cgacgcgcga cttcgagccg gtcgtggcca cggcggtggc ggcgcgagg ctgtatgate	27180
gcatgcgtgc ctgcgcgcgc gtcatgtcg ccaggtagtc gacgggtgc cccgaaatca	27240
ggcggtactc accctgcggg gtcgagaaca cgcgcgaaggaa gcttcaggc cgggtcgta	27300
cgatgcgcgcg gtgggtgtat ccgtcggtt tgctcttcgc gatccacgac gtgtgtccgt	27360
cgccgtggc ctggagcaac tgcgaaagacat acgagtgccg caccgggtt gggagcgaca	27420
gctcgacccgc ggagcgccga ctcaggcat ggaggggccgc ggcgttgaag cgcacccgtt	27480
gccgggcaac ggcgcgtcccc ggcgcggca tcgcggccgc ggggtcgagg accagatct	27540
ggggaaaccgc ggcgaaggcc acggaagaga agacgatcgc cagcgcgcgg aggaaggccg	27600
ctgtatgtcg ttgcgtgaat ggcgttcatgt ggagctcggt aagggtaccc atgcccatac	27660
cgctatgtcg cggagatgtg gacaaaaggat atcgagccgg cgtgcacgacc cgcgcgcgg	27720
gggatgtcc aaaaggacta cgagggtgct acggctgggt tggcgagg aaccaccgac caccacccgg	27780
gaacttcgtt tggctccgc ggcacttaag gtcgaggaa aaccaccgac atcgtcacgc	27840
ggccgatcaa gtccgcgtc cgcgtcggt tgacggccgtt catcaggcgat atcgtcacgc	27900
gcccatcggtt gtgggtggaa accacgtcca tgcgcgcgtt ggcgttgaag tcgaggagct	27960
gcgtcaccgtt ccagcccgcc aacggcgccg ggtatgtcgcc ggcgcgcagg atcgcgggtgc	28020
cgttcatcgat ggcacgtgc ggcgcggcgt cgggtgcgcg gacacgaggat tccgcggccgc	28080
cgatcgccgtt gagatctccc accagggtca cgcgcgcgtt ggcgcgcgg ggcacacgat	28140
cgatcgccgtt ggcgcgcgtt gtcgcacgtt tcaggaaacag gtgtgcgcgg cgcgtcggtt	28200
ggcgaaagac gaggtcgccc ttctgtcgcc cgctgtggat cgcgcgttgc ctcacccccc	28260
agccgcgtgcc gggcgccgagg aagccacgc cggcggtgt cgcgcgttgc ttcatgtatgt	28320
atgcgtatgcc ggcgcgcgtt acgttggaga agacgatgtc ggcttgcgcg tgcgcgttca	28380
tgcgcgcgtt ggcgcaccacg ttccatcccc agcccgccgg cagcaggctc gcaactgcgc	28440
tgatcgccgtt gccattcatc agccagatgt gggcggtcc gtcgtgtgc tgagcagcg	28500
gtcggccctt cgcgtcgccgt tcaggatcgcc cgtgtggctg atgttccagc cgcgtccggc	28560
ggggaaacaggc tccttgcggc ccacgaccgt gagccgcgtt atctggatcg tgcgtatgt	28620
cccgatcggtt tgctggaaat agatgtcgcc ctggcggtcg ccgttggaaat cgcgcgcgtt	28680
tgatcgcgtt ccagccggcg cggcgccggaa tcaggatcgcc cgcgcgttgc atcgtcggtt	28740
cgatcgccgtt ccacgcggcg atgcgcctat cgggtgtgtc gacacggaaat tgcgtctgc	28800
cgatcgccgtt caggatcggtc ggcgcgttgc tcggaaacgc cacccgcgcg aggaaggcat	28860
cgagaaacc cgtgagcaac gtctgtacg cgcgcgcgtt ggtcggtatag tgcgcgcgcg	28920
acgtcgccgcg ggcacatag gccgcgcgcgc ggcgcacatcg gccgcacgcg tacccaccc	28980

cgacgttgtt	gccggccgatc	agcgctcgaga	acacgatgtc	gccgttgcgc	acgcgagtga	29040
ggaacgcgtc	ggcatcgccg	cgggggccct	gcaccgggtt	cagggtcgaa	aagccggcga	29100
cctgcgtgtc	gcccacgacg	agcgcatcg	cattcgccgc	cagggcgacg	tccaggggat	29160
tgtcgtaggc	cgagccgccc	aagatgcgg	acgattcgag	cgccgggttg	gccgagaagc	29220
gcgtcacgaa	cgcacgcacg	gtgcccggca	aattgcgcac	cgttccggtc	tgccccaaagc	29280
ttggcaggcc	ggtatacgcc	acgacggtcg	cctggcccg	gaccgggttc	accgcgatcg	29340
cggtcacgtt	gtcgccgctg	cgcccgagga	acgtgaaata	cgcgaccggg	ccgcccgtcg	29400
caccatctt	ggtgacgaaag	cgcccacccg	gttgggtctg	gtacgcattc	acgttgcgga	29460
acgtcgagga	gcccgtggag	cccgcgacgt	acgtgttgc	cgattgtcc	gttgcgatac	29520
cggggccgat	gtcatcaccg	ctgcccggca	gatacgttga	atagacgaag	ccgcccgttg	29580
acgagagctg	gtgacgaaacg	cgtcgcgatt	gaggccgtg	ccgacgaaca	cggtcgacg	29640
gegctgcgc	cggtcgggaa	gcccgcgcg	gccaacccg	tgacatgc	gaagccgttg	29700
ttgtccacgg	cgatcgccat	gcccctcg	tcgcccggac	cgcacccag	ccgcgagaag	29760
atgacgttgc	cagtggatt	cagcgcgtg	acgaatgc	cgccccgc	ggcatggtg	29820
gtgcaatcgg	tgacggcgaa	gcaatgcggg	aacgcgttgg	tgccggtg	accgggtatc	29880
g gcacgaact	cggtcgcc	ggaggcgccg	agcgcagcc	cgccggagata	cgcttcggcc	29940
ccgagcgtgg	tcgagatcac	cacctgtgc	gtgggtgggt	tgaacttcac	caccgtgg	30000
tcgtacgtgc	ctgcccgtata	ggtcgcgagc	ccgacgtaca	cggtccctg	cgctccacc	30060
ttgaccggcg	tggtccactg	ctcgcccg	ggtacgttga	gaacgcccac	30120	
accgggtcga	tcaccacgt	gccccgtctg	tctgtatcgg	cgatcacaa	gcccgcctcg	30180
gcccctgtct	cgcccacgaa	gagctcgaaag	cgccggccca	ccggcacgcg	cgatcgccg	30240
acctgttgg	aggctaccgg	cgctgtgtc	gtaaaattct	cctggccac	gcccgtcg	30300
aggttgcct	cgccatcgat	ccacgcgtgg	tcggcgacg	agaggtcgag	cgccatctgg	30360
cgcggatcgg	cgcgcccccc	gaccacgaag	tcttattcga	cggtccctc	tttgcgtac	30420
acggtgaggt	ccacgcgcgc	atacaggtcc	ttcagcgaca	cgccggccgaa	gtccggcacg	30480
ttttcgcc	ggccggccggc	cgtggtccg	ctgttaatagt	ggctgagggt	ggccgggggc	30540
tcctcgcc	cgatgaccgg	atcggaaagcc	cccgcaagc	gcacgcgaa	cagcgaggcg	30600
ctggcagctt	tcctggagcg	tggctcgaaag	gcaatgc	cgggctcac	ggagacgcgt	30660
ccggcctgcc	cgcgcgagac	gtagagcgca	tccggcccg	actggccctc	tttgcgtctg	30720
aacgtgatgg	gcccgttcc	gagggcatca	gggacgacgg	catggacggc	ggcgggagtg	30780
gcgagggccg	ccgcgacgag	cacggccagg	ggccggagcg	ccggccgggc	ggccgttggag	30840
ctgggggtca	tggggcagt	acctttctgt	tgttatggct	gttgggttgc	tttgcgggat	30900
ccacagctcg	gatccagctg	gctataactac	cagcgagcga	gtgggttttc	gtgacccagt	30960
cccgcagtcg	ggccccggcg	acaacgcgg	catcgccgtc	tttgc	gcaegcacgg	31020
cctcgatcca	gtcgccgtcg	cgggccagca	cgccgggggc	atcgccgtag	ggctcgagat	31080
ccgaggccac	cacccggatg	ccgaggcagc	ctgactcgag	cagcttcagg	tttgccttc	31140
cgcgattgaa	cgggttgcg	cgcagccgt	ccaccgcgac	gtcgagtgcc	agcgcacgcg	31200
gettcgtccg	gtactcgccg	atgggcacca	tgtcg	ttccgcggcg	aacggccca	31260
gctcgccgt	gcacaggccg	aggaacaccc	atcgatctc	gcatcgctg	gcccgcacca	31320
cgggctcgag	caacttcagg	tctcccat	gtctttcgc	cccccccg	cccacgcgcg	31380
gacgcgcgc	accgcgggaa	cggttcgca	ggccggccca	acgctcgca	tcgatcgctg	31440
tcgggatcac	gcccacgtcc	ttcgcgcgc	ggccgaaaggc	ctccgcgagc	ggtgcgggtgg	31500
acaccacag	gcccgtcgac	aacgcaccc	cgcgcgcgat	gctggggcg	atgtccgggt	31560
agatcgctcg	cgacatacgaa	ttgcccgg	gtcgatcg	gagcaggta	tccaggccca	31620
ggaccttcag	cgccgcgtcc	tggcg	gtacactcgag	cgaggtgagc	ttgttagttgt	31680
ggaagaagtt	gtgcgcgagc	acggcg	cgatcgaggcg	ctgcccactcg	acgcgggttgg	31740
gcgcgcagcc	cgtggcggtc	tgc	cgatcgatcc	gacgcggccc	gcccgtcc	31800
gcgcgcgc	cggtcgccg	acgcgcac	cgccggaccc	cgccgtcg	aaacggaaacg	31860
cgcagagcc	aaccgcgtc	gtgtcg	cgccggagcc	ccagcggtcg	acggaaacg	31920
agccctcgcc	gcccacgc	aacgcgggt	gttgcgtcg	acggggccct	acgcggccgt	31980
gccagcgctc	gcccaccc	gccttc	aaaggatgac	gatcgatcg	accgcggct	32040
cgccatcgcc	gcgcacatcg	cgcgac	ggaaagtccgc	cgcgatcg	ggccgcgc	32100
ggatgtcgaa	gcccgcgtcg	cgccagcc	ggcacagggt	cgccatcg	aacggccccc	32160
cgcgcgtcgat	ctcgatcg	gcctcgccg	agagcgccgc	gtcgccgttgc	accagcgca	32220
gacggcccg	cgcgacgc	acctcgccg	gcatcgatcg	gatgtcccg	aggcgctcc	32280
cgccctcgcc	gcccggccg	acgcgtcg	ggatcgatcg	ggccgcgc	ggccgcgc	32340
cgccctcgcc	gcccggccg	gcccacgaa	cgcccccacgg	ccgcggccccc	gcatggatccc	32400
agccgggaaac	gcccaccc	cgggggccca	cgaggtcg	cgcgagg	cccggtggagg	

ggcctgcgtat	gccctgctcg	aggcgctcg	gccagccggc	gcggaattcg	cgtcaacgcg	32460
catcgacgt	cgcgatccac	agcgtcggc	cgttgcggat	ggcgccgcgc	aacgcggagc	32520
cggtttcgcc	gcccacccgg	accaggtcg	tgcgcacgco	gcgcggcccc	agctcacccgg	32580
gccagegtt	cggcggccct	cgcgctcgaga	tgcacgtaca	cgtggccgat	acgttgcgg	32640
cagcccttca	cgagcgccct	gatgcacgco	accagccgct	ccggcgctg	cgtcgctca	32700
atgaaggcgc	aaatcggtcc	ggccccgatc	gwgcgccccc	acgatgcggg	cgcacatggac	32760
gcagctggca	gaaacgcggc	ctgcgtgc	gtgcggcgga	agtgcctccgc	gaccacgcgc	32820
tgggtggcgt	cgcgccggc	gtcgcttcc	tgcgcggcga	tggtcagggtc	gaggcgatgc	32880
acgagcact	cgcgctgtc	caccacggcg	ctgcacggc	cctccctccgc	gaggcgca	32940
gcgagatcg	tgacgccccg	ccagccggcg	gwgcgccccc	cgccacgctc	cagcaacgc	33000
gcgcccgcg	gcacccgcg	cccgccgagc	ggctgggtc	gcgcgaactc	cgcgcgcgaa	33060
gtggggacg	acgcggatc	gtagcggtc	ccgtgcgc	tcaggcata	actgcagccg	33120
tagagaagc	cggcgttgc	gttgcgtcg	accgcaccccg	ccaccgcgt	cggcgccca	33180
tcggccagcc	ggtgcggccct	ctccaccagc	acgacccatt	cgcccttcgc	gtccgcgatc	33240
gwgccgttga	cgcgcccgag	cgcccgctcg	ctcgcgtcg	gtcgaggac	cacggtctcc	33300
aggcgatcg	cggcgtgtc	cctggccggc	aaggcctcg	gcgcacatcca	gtggctgtag	33360
gcatcgccgc	ccagcgccgt	gwgctcgcc	cggcggtggaa	agccgaggca	cggcgccg	33420
aacaacggat	cgatcggtc	gggcacagca	tccaccgc	catcgccgtc	cggcgccagc	33480
tcgcgataga	gwgccaggta	ctcgccgc	ttcgccctcg	tcgttccgt	gwggtggtag	33540
ttcagggtct	ccgcgacgc	atcgagcgcc	gtgcggtcgt	cgcgaggcg	caggcaggagg	33600
gccacgatcg	cgccgggatc	gttgcggc	agcaggaagc	cggttgcgc	atcgccacg	33660
cgtctggaa	tggcgccatc	gwgctgtccc	gwgccacggg	tgccgaggcg	ctgtgcctcg	33720
ctcaacacca	ggcagaaggt	ctcctccatc	gtcgaggggc	gcaggaccag	gtcgatgc	33780
cgaaggatgc	gwgccaggc	cgtcgccc	taggcgcgc	gtcgatcgat	gcggccgc	33840
ttcgccggcg	cttcgagcgc	gagatcgcc	ggaccccccagg	cctcgatcg	gtcgctgc	33900
cccgccagcc	ggcgccggc	ctcgatgatc	aggtgcgc	ccttgcgc	gcgcagccg	33960
ccgaggaagg	ccacgcgc	caccggccgc	gcttcggggc	gwggaagcag	cggcaggctcg	34020
gcaatgcgt	gcccgtatc	gwgcaatc	gcttcgcgc	tcgcgcgc	gccttgcgc	34080
atgcgcatcg	ccgcgtactc	ggacgggc	aggatcacgt	cgccgcgc	cacggcatcg	34140
gacgcccgg	cgaaagcgc	ttcgaggtag	ccggtcacgg	acgcggcgc	cgtcggcccc	34200
acgatcgct	cgatcgccgc	gwgcaaggca	tccacgc	cgccatcg	gwgccgcgc	34260
gcggcgccgc	cgacacggc	gtcctgcgc	ccgttgcata	tgtttagtc	ggcgacacagg	34320
tacgagaagt	cgatcgccgt	gatgaccac	cgtgcgcgc	actcgcgc	gatgcgcggc	34380
aggcgccaggc	tgttccagcc	caccaac	tggaaatgc	cgacgtcg	gcggccggac	34440
gwgacgtct	tgcgcagc	gttcaccagc	gwgccgtgt	gttcctcaat	gwgccgcgc	34500
gatcgagggc	gtatcgcc	cgatcgac	cgccggccgc	ccgcgcgc	gtgcccgc	34560
gwgccggcgt	tgcgcgc	ggcggccgc	cgacgcgggt	tcgcgtcg	gtgcgtcgcc	34620
acgcccgcga	cgatcggtgg	ccagccgc	tcgaggcc	cgatcgac	cgatcgatcg	34680
gwgctcgatc	cgccaccc	gwgcaac	gcctcgcc	cggccgc	gtggccgc	34740
tcgcagc	cgatctgc	gtcccggtcc	agcgccgc	ccggccatcg	ctcgcgatcg	34800
ccgtatccgc	ggccatagag	cgatcgaa	ggccgcatt	gcacccac	ctcgccgc	34860
accagcatgc	agaagccgac	cgccgtcg	aggcgccg	actcgccgc	gtggccgc	34920
gwgacgcagg	cgagatggc	atcgacgtc	ggctcgcc	cggccgc	gaagatatcg	34980
gwgacgcgg	gwgaggctg	gatcg	cggttggaga	gwgccgcac	gatgcgc	35040
cgcgatcgc	tgcgcggc	acgc	ccctcgagc	aaccggggcc	gacctgcgt	35100
tccgaattga	gacccac	gtcgccgc	ctcgccgc	ggccgcgc	gacgcgc	35160
gwgaaaggc	ggttctcg	attggtgac	acacgatgc	cgccacgc	cgccgcgaa	35220
ctgcccagca	accggcg	gwgccgatc	gtgtcg	catcgac	gacgtgc	35280
tgcgcggc	ccgtgtgg	ggccagc	gwgaggc	gttcgggtgc	gtcgatgc	35340
ttgaagacc	ggacgac	tgcgc	ccatccgc	cgccgttc	agctggcc	35400
atcgccgc	caggcg	tgcgc	agatgc	gatgc	tgcgaggc	35460
cgttaggc	tcatgc	cccagg	tgcgc	ctcgagctc	cgcagcact	35520
gctcg	ggccgc	acccg	gwgcc	gtacatcg	cgccgc	35580
ccgcgc	cgccgc	tcgcgc	ggccg	tcg	cgccgc	35640

ccggccgcggg	tttcgcagcc	tccggcgccg	aagcttgcag	caactgcacg	gcgtcgggca	35880
ccgcgtgcag	cggccagatc	gccgaatacg	catcgacgcg	ttgcccgaac	agggccacct	35940
ccgggaaccg	cgccttcagc	acggcgagga	attcgtgctc	gttagagctcg	cgcacgtat	36000
gcgggttgcg	gtagtcgcgc	tggtcgaggt	acacctcgcg	gttcggcgtc	gagaccagca	36060
gcaggccgccc	gggcgcccagc	acgcgcattcg	cttcgtcgag	caggcgctcg	ggatcgggaa	36120
tgtgctcgag	cgtctcgaaag	gagacgagga	gatcgatgt	tgcgtcatcg	cacggcagcg	36180
actcgcacgc	gcgcctcgacg	tactgcagg	tctgcgcggc	accatagcg	cggcgcgcct	36240
gcgcgttgtt	ctcggcggcg	acatccgcgc	ccaccacgga	cttcgcgcgc	gtcgccagca	36300
gcgcgcgagcc	gtagccctcc	ccgcaggcga	tgtcgaggac	gcggcagcct	ccggcgaagcg	36360
gcaacgcaaa	gtggtagcg	tgccagtgt	cgtaccagat	ctcgccctcg	aagccggct	36420
ggaaaacgttc	gtgttccatc	gggagagtat	aggtgttatgt	gcaaccgcgc	tccggagagg	36480
gcgggttgcga	tgtatcttatt	cgagatcg	tgctacggaa	gcacccggcaa	cggctcgccg	36540
ggcgccaccg	accatccgg	gccccgcgc	agcagggtcg	cgctgcgcag	cgtggtgagg	36600
ccgtccatca	ggcgcacccgt	gatgcggcca	tgcacattct	tgaacaccat	gtcgagcttgc	36660
ccgtcgccat	cgaagtcgag	cagttcg	acgggtccacg	ccgcggccgc	cggcagcacg	36720
tccgcgcgc	tgaggatggc	ggtgcgttc	atcaggcgca	cgtgcgcacg	ccgcgtcggt	36780
tggcggaaaca	cgtatgtcacc	gcccgcgtcg	cggttcatgt	cgcccacgtg	gctcacgatc	36840
cagccggtgc	ccgcgcgcag	gagctcg	ccggcgccgc	acgcggtgcc	gttcatgtatg	36900
aagagatgcg	cgcggccatc	ggtgtggcg	aagaacatgt	ccgccttgc	gtcgccgctc	36960
acgtcgccca	ggtgctcac	cgtccagccg	ctggcgccgc	agaggaagcc	cgcgcggcg	37020
gtgatggtgg	tgcgttcat	cagtagatg	tagccgcgc	cgtcagcg	gatgaagacg	37080
atgtcgccc	tgcgtcgcc	gttgggttcc	ccgggtggca	cgactctcca	gccccgtcgcc	37140
ggcccccagca	gctgctggct	gccgatgtatg	gccgtgc	ccatcagcca	cagggtcgag	37200
cggccgtcg	cgttgcgcag	caggagatcc	gccttgcgt	ccgcgttcat	gtcgccggtg	37260
cggtcgatgc	tccagccgag	gccggcgccc	agcagctct	tgcggccgt	caccgtgagg	37320
ccgttcatcg	tgtacacgt	cacgcgc	tcgggtgtgg	gggatcctct	agagtcgacc	37380
tgcaggcatg	caagcttgcag	tattctatag	tctcacctaa	atagcttgc	gtaatcatgg	37440
tcatactgt	ttcctgtgt	aaattgttat	ccgctcacaa	ttcacacacaac	atacgagccg	37500
aaagcat						37507



**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- BLACK BORDERS**
- IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- FADED TEXT OR DRAWING**
- BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- SKEWED/SLANTED IMAGES**
- COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- GRAY SCALE DOCUMENTS**
- LINES OR MARKS ON ORIGINAL DOCUMENT**
- REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.